

LINGUISTIC ANALYZER: AUTOMATIC TRANSFORMATION OF NATURAL LANGUAGE TEXTS INTO INFORMATION DATA STRUCTURE

Kuznetsov Sergey A.,
Skrebtsova Tatiana G.,
Suvorov Sergey G.,
Klementyeva Anna V.

Reviewers:

L. L. Iomdin (PhD in Philology Sciences, senior scientific researcher of RAS, Institute of Information Transmission Problems named after A.A. Kharkevich, RAS);

D. A. Kocharov (PhD in Philology Sciences, Associate Professor, Saint-Petersburg University Press).

This work was done under the project “Automatic Transformation of Natural Language Texts into Information Data Structure” as part of the event “Conducting fundamental scientific research on the priority areas of the St. Petersburg State University Development Program”

This book was recommended for publication by the Scientific Commission in Languages and Literature Sciences of St. Petersburg State University.

© Saint-Petersburg State University, 2019

St. Petersburg University Press, 2019

ISBN 978-5-288-05927-8

<https://doi.org/10.21638/11701/9785288059278>

The book offers a novel approach for analyzing natural language texts and representing relevant information in a metalinguistic database structure. Methodological principles underlying the processing techniques are the same at all stages. They draw on an empirically-based theory providing for the well-groundedness and consistency of the rules and preventing any possible subjectivity. The ultimate goal of the processing is a conceptual structure made up of elements referring to extralinguistic objects and situations. The elements are assigned ontological properties and linked together by conceptual relations. This makes it possible to extract “pure” information from the text, its reliability being dependent solely on the author

Keywords: computational linguistics, natural language processing, morphological analysis, syntactic analysis, ambiguity resolution, information databases.

CONTENTS

Introduction	5
Chapter 1. Graphematical analysis	16
Chapter 2. Token attribution	27
Morphological analysis of letter tokens	27
Morphological prediction algorithm	32
Attribution of other token types	38
Chapter 3. Constructional syntactic analysis	44
Semantic-syntactic dictionary	44
The notion of semantic-syntactic variant	44
From morphological attribution to semantic-syntactic variant	47
Realization conditions	49
Meaning proper	55
Constructional syntactic analysis algorithm	67
Chapter 4. Communicative syntactic analysis	91
Theoretical assumptions	91
General overview of text information structure	95
From syntactic structures to text information representation	97
Information accumulation: control of objects and features	102
Space and time representation	107
Revealing temporal structure of text	113
Automated recognition of text type	135
The notion of text type	135
Academic writing	139
Legal documents style	142
Journalistic writing	144
Descriptive, narrative, and persuasive writing styles	146
Distinguishing between dictum and modus	163
Chapter 5. Functional-pragmatic text analysis	167
Identification of propositions and their analysis	167
Conceptual types of propositions	176
Fact	175
Opinion	183
Evaluation	191
Expression of will	194
Composite utterance as a unit of rhetorical structure	196
Synthesis of conceptual rhetorical text structure	200
Conclusion	205
Sources	207
Dictionaries and reference books	208
References	209

ADVANCED CONTENTS

Chapter 1. Graphemata analysis

<https://doi.org/10.21638/11701/9785288059278.02>

Graphemata analysis marks the first stage of text processing. However, prior to it, basic text structuring takes place, resulting in the identification of paragraphs and their types, e.g. title, subtitle, author name(s), chapter and section titles, footnotes, endnotes, figures, appendices, epigraphs, etc. After that, graphemata analysis proper begins. Its aim is to decompose the flow of letter and non-letter graphemes into character strings such as individual words, abbreviations, numbers, and hybrid strings (e.g. mathematical formulae). The procedure implies an iterative process of unit assembling, from individual characters to what is called atoms, next to tokens (roughly equivalent to word occurrences), sentence parts and finally, a whole sentence. At every stage, each unit is assigned its type. Assembling relies on the rules based solely on a thorough structural analysis of context. No formal models or statistical methods are applied, this being a central principle of the linguistic analyzer, inherent in all its algorithms. At this stage, complications arise primarily through the ambiguity of punctuation marks. They are discussed at length throughout the chapter.

Keywords: grapheme, letter grapheme, non-letter grapheme, character string, punctuation mark.

Chapter 2. Token attribution

<https://doi.org/10.21638/11701/9785288059278.03>

The aim of this step is to check the previously assigned token types and to provide the letter tokens with morphological information, i.e. values of the relevant grammatical categories. As far as word forms are concerned, the procedure is normally called morphological analysis. However, since a text may contain other token types (e.g. abbreviations, formulae, Internet hyperlinks, phone numbers), it is generally referred to as token attribution. In the chapter, a wide range of token types are considered from the processing viewpoint. In particular, the letter token analysis presupposes a search in a number of dictionaries. Apart from a regular Russian morphological dictionary, search is also performed in the dictionaries of abbreviations, of fixed phrases, and of proper names (personal, geographical, etc.). Morphological analysis often yields more than a single attribution. In some cases, the ambiguity can be reduced by taking into account graphemata information, but most often, it will remain and pose further problems for the syntactic analysis. If the search for a letter token fails in all the dictionaries, the algorithm tries to identify its lemma and predict the grammatical meaning. The attribution of other token types is made by mapping them onto a range of patterns. Typical problems of both operations are discussed.

Keywords: morphological analysis, tokenization, abbreviation, fixed phrase, proper name, morphological prediction.

Chapter 3. Constructional syntactic analysis

<https://doi.org/10.21638/11701/9785288059278.04>

At this stage, a semantic-syntactic dictionary is activated, in order to supply the letter tokens with the information concerning their contextual realization and meaning proper. The list of features assigned to a word form is thus augmented by semantic and syntactic ones. Contextual realization is specified by means of subcategorization frames. Meaning is described in terms of semantic features which may or may not be hierarchically arranged. Among them, basic features are identified, e.g. 'living being', 'action', 'space', 'time', 'quantity', that have no direct bearing on the parts of speech. Thus, the feature 'action' can be assigned both to verbs and nouns. Additional semantic features make it possible to account for manifold aspects of word meaning. Thus, horse is marked both as 'domestic animal' and 'vehicle', which would be impossible in case of a strict hierarchy of features. The algorithm of syntactic analysis draws on the same basic idea as that of graphematical analysis, viz. iterative assembling of individual tokens into syntactic constructions, from local groups to larger ones until a resulting syntactic structure of a whole sentence emerges consisting of Subject-Predicate-Objects-Adverbials. Syntactic rules are ranked accordingly. The chapter illustrates the work of algorithm in a variety of cases and discusses multiple difficulties resulting from the syntactic ambiguity.

Keywords: semantic feature, subcategorization frame, syntactic analysis, syntactic rule, syntactic ambiguity.

Chapter 4. Communicative syntactic analysis

<https://doi.org/10.21638/11701/9785288059278.05>

The goal of communicative analysis is to transform the Subject-Predicate-Objects-Adverbials structure of individual sentences into a metalinguistic information structure. In the latter, two basic data types are identified, objects and attributes. Attributes are linked to a single object or establish connections between a number of objects (in which case they can be also called predicates). An attribute together with its object(s) forms a proposition. In the information structure, objects can be related to one another either directly or via attributes. Attributes are also interrelated by temporal, causal, and other links. There is no direct correspondence between the constituents of the Subject-Predicate-Objects-Adverbials structure and the information structure components, e.g. the subject of a sentence does not necessarily become an object in the information structure, etc. The analysis of successive sentences requires to keep track of objects and attributes, which brings in the problem of reference. Information on space and time should also be accounted for and represented in the overall structure. As no restrictions are imposed on the type of the analyzed texts, the chapter summarizes the distinctive features of various text types. However, they prove to be insufficient for the type identification. The situation is further complicated by their intricate combinations (narration can alternate with description, persuasion can enter into narration, etc.). This chapter is a pilot study of quite a number of issues central for natural language processing but as yet lacking attention and satisfactory solution.

Keywords: information structure, proposition, reference, temporal structure, text type.

Chapter 5. Functional-pragmatic text analysis

<https://doi.org/10.21638/11701/9785288059278.06>

Drawing on Charles Bally's distinction between dictum and modus, the final chapter deals with the pragmatic analysis of propositions identified at the previous step. Propositions can be expressed by an attribute linking together a number of objects or by a single word (implicit proposition). Implicit propositions should be made explicit, i.e. missing objects should be brought out. Each proposition is attributed a conceptual type. The authors posit four types of propositions, namely, fact, opinion, evaluation and expression of will. While the distinction between fact and opinion is common enough, evaluation emphasizes the emotional, or expressive, aspect of an utterance, and the expression of will has to do with persuasive speech acts. Next, relations between propositions, both local and distant, are established, which makes it possible to bring them together within a rhetorical text structure. The units of rhetorical structure are quite different from both the syntactic and communicative ones and vary depending on the its type (statement, persuasion, directive, etc.).

Keywords: proposition, fact, opinion, evaluation, expression of will, rhetorical structure

FOREWORD

<https://doi.org/10.21638/11701/9785288059278.01>

Over the last decades, the challenging problem of natural language processing has received much attention from scholars of linguistics and computer science. However, it still remains a topical issue. Despite recent advancements in particular areas, the development of new tools and techniques, few frameworks have been created accounting for the overall processing of text and involving all types of linguistic analysis.

The book outlines an original approach for the automated analysis of Russian texts aimed at revealing their information structure. The linguistic analyzer devised by the authors begins with an unstructured Russian text and, after performing graphematical, morphological, syntactic, and communicative analyses, transforms it into a metalinguistic network. No restrictions on the text type are imposed, except that it should be written in the standard Russian language.

The authors present a pilot study and do not aim to give a full account of the linguistic analyzer. The focus is primarily on the linguistic aspects of its work rather than on the particular algorithms. Furthermore, they put forward certain fundamental principles that cut across text processing at different stages and make the proposed approach consistent and well-integrated. These principles constitute the theoretical background of the book and are stated below.

As has already been mentioned, the analyzer is supposed to operate on various text types, regardless of their style and genre, and process them from the lowest level of individual characters up to the whole discourse. Normally, natural language processing is limited to a particular text type, task, or kind of analysis. Thus, the proposed approach is much wider in scope than the majority of projects in the field.

A central feature of the proposed approach is the principal idea of relying solely on the linguistic information contained in text. The analyzer does not resort to probability calculations or statistical methods. While admitting the usefulness of mathematical apparatus for the narrow-scope linguistic applications, the authors believe that a comprehensive interpretation of text requires qualitative rather than quantitative techniques. Thus, the emphasis is placed on a thorough linguistic analysis. After all, human beings do not have to calculate probability in order to understand natural-language texts.

The work of the linguistic analyzer encompasses a number of stages roughly corresponding to the book chapters. It begins with graphematical analysis of individual characters, and proceeds to morphological, semantic, syntactic, and communicative analyses. The latter includes quite a number of text organization aspects, including those of pragmatics and rhetorical structure. A major focus is on the constructional syntactic analysis. In the processing, a number of dictionaries are activated, either compiled from scratch or adjusted and augmented by the authors. In particular, there are dictionaries of graphemes, abbreviations, fixed phrases as well as morphological and semantic-syntactic dictionary.

Linguistic analysis consists in the interpretation of information contained in the verbal component of the text. As a result, units are assigned values of relevant features. This is done in a strictly progressive manner, beginning with the lowest-rank graphematical characters up to the highest-rank syntactic groupings, viz. group of subject, group of predicate, object groups and those of adverbials. Characters (graphemes) are specified in terms of their type (with the values letter / non-letter character, the former being further characterized as belonging to a particular writing system and the latter being subdivided into number / non-number characters). Morphological analyzer assigns word forms values of grammatical categories relevant for the Russian language (number: singular / plural, genus: masculine / feminine / neuter, aspect: perfect / imperfect, tense: past / present / future, etc.). At the stage of syntactic analysis, values of semantic features are added on the basis of semantic-syntactic dictionary as well as values of syntactic ones, specifying the role of the word or phrase in the sentence. Thus, information extracted at every processing stage is uniformly expressed by tags capturing features and their values.

It should be emphasized that very little useful information concerning relevant features and their values can be automatically extracted from the existing dictionaries and reference books, as they are not oriented towards natural language processing. The algorithm uses rules of if... then... format, and no traditional books on language present information in this way. So, much empirical work had to be done to both re-formulate the existing knowledge and bring out more specific regularities. Russian National Corpus containing vast amount of current usage was a valuable resource the authors used for that purpose. The whole procedure proved to be effort- and time-consuming.

The book discusses at length classical problems of computational linguistics from the viewpoint of the analyzing algorithms and illustrates them with samples found in a variety of fictional, mass-media, legal, etc. texts. Among such problems, ambiguity resolution stands out most sharply, plaguing all natural-language processing endeavors at all levels of analysis. With respect to the analyzer concerned, ambiguity resolution would mean picking up a single interpretation of feature values out of a whole range of options. Unlike some other computational models, the analyzer is not supposed to cope with ambiguity as soon as possible, at the very stage when it occurs. Rather, ambiguity resolution is postponed till the moment when it can be safely dealt with. Thus, morphological ambiguity is usually cancelled at the syntactic stage while reference resolution would require going beyond sentence boundaries.

Along with practical computational issues, the book touches upon some global and “eternal” problems, such as What is word? and What is sentence? This was not originally among the authors’ goals and certainly does not constitute a priority. Rather, it is a by-product of working with real language data and attempting to map them onto the rigid traditional concepts.

The ultimate goal of the processing is a metalinguistic network, with objects being assigned certain properties and linked together by predicates. Such network could be used in a wide range of applications bearing on the collection, updating, search, and extraction

of information. However, the very transition from a syntactic structure to a metalinguistic network poses a host of problems, being a largely untrodden ground. It is here that the exploratory character of the study reaches its climax. The authors hope that the discussion in the book will open up new perspectives on the extension of natural language processing beyond the sentence level.

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ЛИНГВИСТИЧЕСКИЙ АНАЛИЗАТОР

Преобразование текста
в метаязыковую
структуру данных



ИЗДАТЕЛЬСТВО САНКТ-ПЕТЕРБУРГСКОГО УНИВЕРСИТЕТА

УДК 81'32

ББК 81.1

Л59

Рецензенты: канд. филол. наук., вед. науч. сотр., Л. Л. Иомдин (ИППИ РАН им. А. А. Харкевича), канд. филол. наук, доц. Д. А. Кочаров (С.-Петерб. гос. ун-т)

Рекомендовано к печати

научной комиссией в области наук о языках и литературе

Санкт-Петербургского государственного университета

Лингвистический анализатор: преобразование текста
Л59 в метаязыковую структуру данных / С. А. Кузнецов, Т. Г. Скребцова, С. Г. Суворов, А. В. Клементьева. — СПб.: Изд-во С.-Петерб. ун-та, 2019. — 238 с.
ISBN 978–5–288–05927–8

В пяти главах монографии, упорядоченных согласно последовательности этапов обработки текста, рассматриваются графематический анализ, морфологический (атрибуция токенов), конструктивно-синтаксический и коммуникативно-синтаксический анализ предложения, функционально-прагматический анализ текста. Продемонстрирована работа лингвистического анализатора, процесс преобразования понятийно-риторических единиц текста и информационного представления текста в метаязыковую структуру данных. В приложениях даны анализ предложений из повести Ю. В. Трифонова «Обмен», примеры статей семантико-синтаксического словаря, примеры семантического разбора законодательных текстов.

Книга предназначена для специалистов в области автоматической обработки текста и будет полезна в практике преподавания компьютерной лингвистики.

УДК 82.09

ББК 83.3

Работа выполнена по проекту

*«Автоматическое преобразование текстовой информации
в метаязыковую структуру данных» в рамках мероприятия*

*«Проведение фундаментальных научных исследований
по приоритетным направлениям Программы развития СПбГУ»*

© Санкт-Петербургский
государственный
университет, 2019

ISBN 978–5–288–05927–8

© Авторы, 2019

ОГЛАВЛЕНИЕ

Введение	5
Глава 1. Графематический анализ	16
Глава 2. Атрибуция токенов.....	27
Морфологический анализ буквенных токенов	—
Алгоритм морфологического предсказания	32
Атрибуция других типов токенов	38
Глава 3. Конструктивно-синтаксический анализ предложения	44
Семантико-синтаксический словарь	—
<i>Понятие семантико-синтаксического варианта.....</i>	—
<i>От морфологической атрибуции словоформы к ее</i> <i>семантико-семантическому варианту.....</i>	47
<i>Условия реализации</i>	49
<i>Собственное значение.....</i>	55
Алгоритм конструктивно-синтаксического анализа.....	67
Глава 4. Коммуникативно-синтаксический анализ	91
Теоретические предпосылки	—
Общая концепция информационной структуры текста.....	95
От синтаксических структур к информационному представлению текста.....	97
Накопление информации: контроль за объектами и признаками	102
Представление информации о пространстве и времени.....	107
Определение временной структуры текста.....	113
Автоматическое распознавание типа текста.....	135
<i>Понятие типа текста</i>	—
<i>Научный стиль</i>	139
<i>Официально-деловой стиль.....</i>	142
<i>Публицистический стиль</i>	144
<i>Функционально-смысловые типы речи: описание,</i> <i>повествование, рассуждение.....</i>	146
Разграничение диктума и модуса.....	163

Глава 5. Функционально-прагматический анализ текста	167
Выявление пропозиций и их анализ	—
Функционально-понятийные типы пропозиций	176
<i>Факт</i>	—
<i>Мнение</i>	183
<i>Оценка</i>	191
<i>Волеизъявление</i>	194
Композитное высказывание как единица риторической структуры текста	196
Синтез понятийно-риторической структуры текста	200
Заключение	205
Источники	207
Словари и справочники	208
Литература	209
Список сокращений в примерах	221
Приложение 1. Анализ предложений из повести Ю. В. Трифонова «Обмен»	222
Приложение 2. Примеры статей семантико-синтаксического словаря ...	228
Приложение 3. Примеры семантического разбора законодательных текстов	235

ВВЕДЕНИЕ

Проблема автоматической обработки естественного языка сохраняет свою актуальность по прошествии более полувека со времени первых опытов в данном направлении. Сам факт говорит о высокой сложности задачи и отсутствии ясного представления о путях ее решения. В научной литературе последних десятилетий нашли отражение усилия многих исследовательских коллективов по моделированию частных аспектов анализа и синтеза естественно-языковых текстов, предложены различные методики и приемы, позволяющие добиваться приемлемых результатов для конкретных узких задач. В то же время можно буквально по пальцам пересчитать системы, предполагающие полный автоматический «проход» от текста к его смыслу и/или наоборот.

В настоящей монографии предлагается оригинальный подход к решению задачи автоматического анализа текста, написанного на русском языке, направленный на выявление его информационной структуры. Конечным результатом анализа является метаязыковая сеть, в узлах которой находятся «участники» описываемой в тексте ситуации (объекты), а на дугах — отношения между ними. Программное обеспечение, реализующее данный подход, охватывает все этапы обработки текста на естественном языке (ЕЯ) — от графематики до коммуникативного синтаксиса целого текста. В качестве материала анализа могут использоваться любые правильно построенные письменные тексты публицистического, научного, официально-делового стиля, а также художественной прозы. Это дает нам основание называть нашу модель лингвистическим процессором, понимая под этим термином то же, что и в монографии «Лингвистический процессор для сложных информационных систем» [Апресян и др. 1992: 3].

Книга носит поисковый характер, что отчетливо прослеживается в содержании. Авторы не ставили перед собой цели подробного описания лингвистической модели, включая исчерпывающий перечень правил, применяемых на каждом этапе анализа¹. Основную ценность, на их взгляд, представляют фундаментальные принципы, на которых основывается предлагаемый подход к автоматической обработки текста. Эти принципы последовательно воплощаются на каждом этапе работы лингвистического процессора. Именно на них стоит остановиться здесь, перед тем как перейти к изложению частных аспектов, связанных с тем или иным видом анализа.

Описываемый подход отличается несколькими существенными чертами, выделяющими его на фоне современных разработок в области автоматической обработки текста. Прежде всего, он, как уже было сказано, является максимально широким, не ограниченным какой бы то ни было узкой задачей, отдельным видом лингвистического анализа или конкретным типом текста. Он носит комплексный характер и включает учет всех видов информации, содержащейся в вербальном компоненте текста на ЕЯ.

Важной отличительной характеристикой разрабатываемого лингвистического процессора является принципиальный отказ от применения вероятностно-статистических методов на этапе анализа текстовой информации. Признавая полезность аппарата математической статистики для ряда узко-прикладных задач в области языкознания, мы считаем, что при создании полномасштабной системы автоматического понимания текста на ЕЯ решающей является не количественная, а качественная сторона. Разнообразные логико-статистические приемы анализа текста, исходящие из примата средств программной реализации и сосредоточенные на поиске все более эффективного алгоритма, который позволил бы хотя бы на долю процента улучшить результат, представляются нам методологически ошибочными. С нашей точки зрения, адекватный анализ содержания может быть достигнут только при тщательном учете языкового контекста, поэтому приоритетной задачей является максимально полное и точное извлечение лингвистических данных из текста. Поскольку речь идет о текстах на русском языке, это объясняет заметное преобладание в библиографии русскоязычных источников.

¹ Описание алгоритмической составляющей модели тем более не входит в задачи книги.

В основе предлагаемого подхода лежат правила интерпретации различного рода информации, заключенной в вербальном компоненте текста. Итогом интерпретации является приписывание анализируемым единицам (символам, словоформам, словосочетаниям, синтаксическим конструкциям, риторическим структурам) значений признаков. Имеются в виду графематические признаки (ср. признак «тип» и его значения ‘буквенный’ / ‘небуквенный’, признак «алфавит» и его значения ‘кириллица’ / ‘латиница’ / ‘прочий’), морфологические признаки, включая часть речи (ср. «род» и его значения ‘мужской’ / ‘женский’ / ‘средний’), синтаксические (например, «тип синтаксической конструкции» со значениями ‘группа сказуемого’, ‘обстоятельственная группа’ и др.) и семантические признаки (ср. признак «фазовость» и его значения: ‘становление’ / ‘возникновение’ / ‘начало’ / ‘осуществление’ / ‘продолжение’ / ‘приостановка’ / ‘возобновление’ / ‘конец’ / ‘предел’ / ‘результат’).

Таким образом, вся извлекаемая из текста графическая, грамматическая и лексико-семантическая информация последовательно представляется в терминах признаков и их значений.

Подобная методологическая последовательность, хотя и делает описание несколько тяжеловесным, представляется нам принципиально необходимой. Следует отметить, что слово *значение* в данной работе используется в двух разных смыслах: как ‘языковое значение’ (ср. англ. *meaning*) и как ‘значение того или иного признака’ (англ. *value*). Что касается слова *признак*, его употребление также неоднозначно. Там, где речь идет о графематическом, морфологическом, лексико-семантическом и синтаксическом анализе, понятие «признак» используется в смысле ‘категории, которой присущи определенные значения’. Однако применительно к организации информационной структуры текста *признак* фактически обозначает информацию, приписываемую в тексте выделенным объектам (один из основных компонентов метаязыковой структуры текста), в том числе и соединяющую их друг с другом.

Переходя к другим особенностям описываемого лингвистического процессора, подчеркнем очевидный факт: традиционные лингвистические описания — словари и грамматики — не ориентированы на задачи автоматической обработки языка, а следовательно, нужные для этой задачи сведения не могут быть получены без дополнительных описаний значений слов и синтаксических структур. В частности, требуется тщательная работа с корпусами текстов, с тем чтобы необходимую информацию можно было выразить им-

пликативно («если... то...»). При отказе от математических методов выработка такого рода эвристик — путь довольно затратный, но способный давать эффективные решения. Мы стремились прежде всего к выработке правил, охватывающих не единичные факты, а их классы, хотя и частные фильтры, касающиеся отдельных значений слов и словосочетаний, представляют известный интерес². На страницах книги читатель найдет немало эмпирически выведенных закономерностей, основанных на кропотливом анализе языковых данных. Собственно говоря, именно такой ракурс анализа текста наиболее точно отражает исследовательский интерес авторов.

Тема неоднозначности языковых выражений и, следовательно, принципов выбора правильной интерпретации проходит через всю книгу. Однако следует подчеркнуть, что, в отличие от подавляющего большинства лингвистов, работающих в области автоматической обработки ЕЯ, мы не ставили перед собой задачу во что бы то ни стало и как можно раньше снимать лексическую и грамматическую неоднозначность. Осознавая значимость этой проблемы, возникающей на всех этапах анализа текста, мы не считаем нужным решать ее сразу и любой ценой. В общем случае неоднозначность сохраняется вплоть до того момента, когда лингвистический анализ позволит ее надежно разрешить. Известно, что лексико-грамматическая омонимия, как правило, снимается на уровне синтаксиса, а для определения референции анафорического выражения нередко требуется выход за границы предложения. В то же время есть случаи, когда неоднозначность не может быть снята в принципе — например, если анафора отсылает к внешнему миру (так называемая «экзофора» [Halliday, Hasan 1976: 18]) или авторский текст намеренно двусмыслен. В подобных ситуациях сохраняются все допустимые варианты прочтения. Следует также добавить, что в задачи лингвистического процессора не входит выявление подтекста, понимание юмора и иносказаний, оценка истинности высказываний, определение художественной ценности текста и т. п.

Работа процессора состоит из нескольких этапов, соответствующих типам анализа (в тексте это отдельные главы). Процессор работает сугубо поступательно. На каждом этапе порождаются все

² Показателен в этом отношении пример из работы [Апресян и др. 1992: 76], посвященный снятию лексико-грамматической омонимии словоформы *механику*, которая в сочетании с предлогом *в* отсылает к лемме *механика*, а с предлогом *к* — к лемме *механик*.

возможные варианты интерпретации соответствующих единиц. Следующий этап может либо отменить какие-то из вариантов, либо сохранить их все. Возврат назад, к предшествующему этапу, для пересмотра результатов анализа не имеет смысла (поскольку все допустимые варианты были своевременно порождены).

Помимо алгоритмов, работа анализатора поддерживается специально созданными словарями, в том числе словарем аббревиатур и сокращений, словарем лексических комплексов (включающих составные предлоги, наречия и пр., а также коллокации и фразеологизмы) и словарем частотных имен собственных. Синтаксический анализ опирается на разработанный авторами семантико-синтаксический словарь, описанию которого посвящен специальный раздел.

Сначала осуществляется графематический анализ всего текста, алгоритм которого является авторской разработкой. Основной результат графематического анализатора состоит в выделении структурных единиц текста (от символа до абзаца) с приписанными им значениями графических признаков. Эти единицы, снабженные соответствующей информацией, используются на последующих этапах работы процессора. Например, такая единица, как токен, далее подвергается морфологической атрибуции. Синтаксический анализ опирается на выделенные ранее части предложения (ограниченные нетерминальными знаками препинания) и предложение. Для построения информационной структуры текста важны как связи между предложениями внутри абзаца, так и межабзацные связи.

После графематического анализа наступает этап атрибуции графических токенов, который в случае словоформ соответствует тому, что традиционно называется морфологическим анализом, хотя в действительности морфоанализаторы предлагают лишь возможные варианты морфологических характеристик словоформы.

Описываемый нами процессор способен обрабатывать более широкий круг токенов, регулярно встречающихся в тексте, а именно аббревиатуры, сокращения и прочие комплексы, которые наряду с буквенными символами могут содержать знаки иной природы. В основе собственно морфологического анализа лежит словарь словоформ, созданный в результате предварительных исследований, связанных с автоматической обработкой текстов, но существенно пополненный в ходе настоящей работы.

В связи с этим нам представляется, что было бы уместно называть этот этап морфографической атрибуцией. Во-первых, на вход поступают различные цепочки символов, в том числе и такие, кото-

рые содержат цифры, знаки препинания и другие символы; таким образом, дело не ограничивается собственно языковым материалом. Во-вторых, даже там, где процессор имеет дело с однородной последовательностью кириллических символов, он фактически производит не анализ, а механическое сопоставление этой последовательности (как некой целостной единицы) с массивом хранящихся в памяти цепочек символов (также целостных единиц) и, если находит совпадения, выдает их все в качестве результата. Это процедура сравнения, а не анализа.

Центральное место в описываемой модели принадлежит синтаксису. Все содержание глав 3 и 4, охватывающее формально-синтаксический анализ, семантическое описание лексики, снятие лексической и синтаксической неоднозначности, а также правила преобразования семантико-синтаксической структуры в информационную модель и сам проект последней, представляет собой исключительно авторские оригинальные разработки.

Принципиальный аспект синтаксического анализа состоит в том, что выявление структуры предложения осуществляется не через построение древесных структур «сверху вниз» (ср. англ. *top-down*), а путем последовательной циклической сборки «снизу вверх» (ср. англ. *bottom-up*), в соответствии с линейным порядком слов. Эта идеология, с одной стороны, отличает наш процессор от различных формальных моделей анализа естественного языка, а с другой — сближает его с современными когнитивными исследованиями, направленными на изучение процессов понимания дискурса.

Закономерно встает вопрос, что следует рассматривать в качестве *предложения*, если сугубо формальный (графический) подход не всегда дает удовлетворительные результаты. Во-первых, существуют неполные предложения, интерпретация которых требует обращения к предшествующему предложению (ср.: *Всегда он наталкивался на твердое сопротивление Лены, и с годами идея стала являться все реже. И то лишь в моменты раздражения* (Ю. Трифонов. Обмен)). Во-вторых, в случае прямой речи имеет смысл учитывать в качестве единого содержательного блока весь ее отрезок, который может состоять из сколь угодно большого числа графически оформленных предложений. Следует подчеркнуть, что в нашей модели предусмотрены возможности обрабатывать оба указанных случая (подробнее см. гл. 1), а также анализировать непроективные предложения (ср.: *Послушай: далеко, далеко на озере Чад изысканный бродит жираф* (Н. Гумилев. Жираф)).

Вообще, в работе по созданию лингвистического процессора всплывает целый ряд «вечных» теоретических проблем языкознания, таких как отдельность и тождество слова (ср. обработку сложных союзов, составных предлогов, наречий, числительных, сказуемых и пр.), композициональность семантики (проблема коллокаций и идиом), лексическая многозначность и способы ее описания и пр. В итоге ключевые термины лингвистического анализа — *слово* и *предложение* — оказываются довольно условными, допускающими вариативность в интерпретации. В своем изложении мы, стараясь сохранять ясность, не употребляем термин *слово* в главе о графематическом анализе, используем термин *словоформа* в морфологическом контексте и оперируем понятием семантико-синтаксический вариант (ССВ) слова там, где речь идет о лексической семантике (термина *лексема* ввиду его неоднозначности также приходится избегать). С термином *предложение* сложнее — заменить его нечем, поэтому мы здесь и позволили себе данное отступление.

Синтаксический анализ производится с учетом семантической информации, источником которой служит специально созданный семантико-синтаксический словарь, направленный на многоаспектное описание лексики в терминах семантических признаков, каждый из которых может иметь два и более значений (в итоге получается нечто наподобие семантических формул в лингвистическом процессоре В. А. Тузова [2003]). В этом заключается существенное отличие нашей модели от других семантических представлений, будь то электронные онтологии или словари (семантические, идеографические), составители которых приписывают фактически одному семантическому признаку многочисленные значения, обычно организованные либо в виде иерархического дерева (например, (Шведова 1998–2007)³), либо в виде общего списка отдельных морфологических, синтаксических, лексических свойств, характеризующих слово целиком или то его значение, которое, по мнению разработчиков, является наиболее частотным (например, [Апресян и др. 1992]).

Основное назначение семантического описания — способствовать снятию лексико-грамматической омонимии или, по крайней мере, сокращению числа вариантов, полученных в результате синтаксического анализа. Прежде всего это касается однозначности структурной схемы предложения. Так, с формальной точки зрения в предложении *Автомобиль сделал поворот* оба существительных

³ Ссылки на словари и справочники даются в круглых скобках.

(*автомобиль и поворот*) могут быть как подлежащим, так и прямым дополнением. Благодаря семантическому описанию слов, составляющих данное предложение (*автомобиль* — транспортное средство, *сделать* — глагол действия, *поворот* — отглагольное существительное со значением движения), удастся отсеять неверный вариант синтаксического разбора. Однако каким бы точным ни было семантическое описание, оно не всегда позволяет получить единственную структуру там, где предложение допускает множественную синтаксическую интерпретацию, ср. знаменитый пример *Мать любит дочь*, а также изречение *Бытие определяет сознание* (неоднозначность которого, впрочем, обычно не обращает на себя внимание). Синтаксическая омонимия, действительно, не такое уж редкое явление (ср. [Мельчук 1974: 31–32; Муравенко 2008: 159–162]).

Семантическое описание позволяет также снижать уровень лексической неоднозначности, а в идеале и вовсе ее снимать. Когда синтаксически связанные слова неоднозначны, алгоритм осуществляет последовательный перебор всех соответствующих комбинаторных вариантов с целью сопоставления приписанных им значений семантических признаков. В основе данной операции лежит известная идея о том, что такие слова не должны иметь противоречащих друг другу сем.

Рассмотрим в качестве примера предложение *Рабочий отстоял смену*. Допустим, что грамматическая омонимия словоформы *рабочий* уже снята, а в качестве существительного она имеет единственное значение. Остается разобраться со словами *отстоять* и *смена*, каждое из которых характеризуется довольно высоким индексом неоднозначности. Согласно словарю (БТС 2009), среди значений глагола *отстоять*⁴ есть такие, которые имеют временной параметр: одно предполагает целенаправленную деятельность человека (*Отстоять два часа в очереди*), а другое — темпорально ограниченное существование предмета (*Храм отстоял пять веков*). А существительное *смена*, наряду с обозначением лица (или их совокупности), предмета (одежды) и выражением абстрактного понятия изменения, может использоваться для обозначения промежутка времени. Таким образом, имеет место согласование по семе времени между одним значением существительного и двумя значениями глагола. Число комбинаторных вариантов интерпретации, таким образом,

⁴ Точнее, речь идет о значениях соответствующих омонимичных глаголов.

сократилось до двух. С учетом семантического типа подлежащего *рабочий* неоднозначность удаётся полностью снять.

Несмотря на большой объем используемого словаря, окказиональные употребления слов в расширительном или образном смысле могут, разумеется, приводить к такому «семантическому рассогласованию» [Пак 1972: 381] внутри отдельного словосочетания, что ни одна из комбинаций значений семантических признаков не дает удовлетворительной интерпретации. Подобная ситуация является важным сигналом, свидетельствующим о необходимости переосмысления одного из членов словосочетания; какого именно — решается на основе анализа последующего текста.

У семантического описания есть и другая важная функция, которая обращена не вспять (к синтаксическому анализу отдельного предложения), а, напротив, вперед — к интерпретации последовательности предложений, вплоть до целого текста. Дело в том, что комбинации значений семантических признаков у лексических единиц, связанных между собой синтаксическими отношениями, обуславливают категориальный тип описываемой ситуации. Например, сочетание кванторного выражения со значением всеобщности с глаголом ментального восприятия (типа *Каждый американец думает, что..., Все считают, что...*) обозначает вовсе не «массовую» интеллектуальную деятельность (что получается из буквального сложения смыслов), а наличие (в какой-то социальной группе) некоторой установки, нормы, верования и т.п., при отсутствии конкретного субъекта — носителя соответствующих представлений⁵. Когда за подобным предложением следует что-нибудь вроде: *Эта идея нашла поддержку/не выдерживает критики/является верной/полезной/тлетворной/и т.п.*, — именно семантическое описание, основанное на корреляции семантических признаков и их значений, способно обеспечить когеренцию текста. В противном случае результат автоматического анализа будет напоминать известную поговорку «В огороде бузина, а в Киеве дядька».

Семантика, таким образом, перебрасывает мостик к следующему этапу — коммуникативно-синтаксическому анализу, распространяющему свое действие на более крупные, чем предложение, отрезки текста. Его организация связана с семантикой еще теснее, чем синтаксис и нижние уровни языковой структуры [Откупщикова 1982:

⁵ Субъект в этих случаях имеет универсальный референциальный статус [Падучева 1985: 95–96].

21], что является существенным осложняющим фактором. Ведь из традиционных компонентов лингвистического процессора — морфологического, синтаксического и семантического — «достаточно высокой, если не исчерпывающей полноты можно добиться лишь в первых двух компонентах» [Апресян и др. 1992: 113].

Заключительный этап анализа носит отчетливо выраженный поисковый характер и потому описан в книге более схематично, чем предыдущие этапы. Большое внимание уделено принципам перехода от полученной на предыдущем этапе синтаксической структуры в виде структуры «подлежащее — сказуемое — дополнение — обстоятельство» (ПСДО) к субъектно-объектной метаязыковой информационной модели. Подробно рассматриваются различные варианты соответствия между компонентами синтаксической и информационной структур и проблемы, возникающие при их соотнесении.

При описании концепции информационной структуры, в которую преобразуется неструктурированный текст на ЕЯ, неизбежно затрагивается широкий круг задач, подлежащих решению на этом этапе (определение объектов и связей между ними, экспликация неявных компонентов смысла, развертывание пропозиций и выявление отношений между ними, определение кореференции именных групп, интерпретация дейксиса, установление хронологической последовательности событий и др.). В качестве материала привлекаются тексты разных стилей и жанров, что призвано наглядно продемонстрировать невозможность выработки единого подхода к указанным проблемам. Заметим, что имеющаяся литература весьма скупа по части их возможного решения. Более того, отсутствуют даже сколь бы то ни было развернутые описания различных типов текстов в соответствующих аспектах. За исключением, возможно, некоторых жанров научного дискурса, а также языка официальных документов, лингвистам очень мало известно о том, как устроены те или иные типы текстов, каковы их лексические и грамматические особенности, что собой представляют временные характеристики содержания, какие языковые средства используются для воплощения соответствующей темпоральной структуры и т. д.

Полезность информационной структуры для решения широкого круга прикладных задач не вызывает сомнений. На ее основе могут быть построены базы знаний, позволяющие хранить сведения (факты, утверждения), относящиеся как к уникальным объектам, имеющим собственное имя (*Петров, Москва, Украина*) или дескрипцию (*президент Молдавии; принадлежащий А. А. Сидорову автомобиль;*

памятник Минину и Пожарскому; день рождения А. В. Суворова), так и к классам объектов (*самолет Ту-134, дизельное топливо, курс валют, биржевые сделки*). Пополнение базы знаний должно происходить автоматически по мере анализа текстовых массивов, получаемых из Интернета или локальной сети. В результате может быть получена информационно-аналитическая система, способная производить мониторинг соответствующих текстов, интегрировать данные из разных документов, осуществлять семантический поиск по объектам, событиям, тематическую рубрикацию и пр.

Помимо этого, возможно применение модели для проверки текста на однозначность — это важно, в частности, для задач, связанных с лингвистической экспертизой.

Глава 1

ГРАФЕМАТИЧЕСКИЙ АНАЛИЗ

Автоматическая обработка текста начинается с предварительного этапа композиционного анализа, направленного на выделение абзацев и определение их типов — заголовков, подзаголовков, имен авторов, названий глав и разделов, сносок, примечаний, приложений, эпиграфов и т. д. Происходит также выделение в составе текста схем, таблиц, рисунков с подрисовочными надписями (ср.: [Леонтьева 2006: 49–50]).

Далее следует собственно графематический анализ, посредством которого осуществляется токенизация (англ. *tokenization*), — разбиение потока алфавитных и внеалфавитных графем на цепочки символов, такие как отдельные слова, аббревиатуры, буквенные сокращения, цифровые последовательности, комплексы смешанной природы (формулы и пр.). Используемый нами подход кардинальным образом отличается от тех, что обычно практикуются в существующих системах автоматической обработки языка¹.

Во-первых, наш графематический анализатор, как и лингвистический процессор в целом, не использует формальные модели и аппарат математической статистики (ср. [Уразлин 2005; Седунов 2007; Ле 2011]). В его основе лежит процедура последовательной сборки все более крупных сегментов текста (символы → атомы → токены → части предложения → предложение), предполагающая тщательный структурный анализ контекста и опирающаяся на разработанные авторами оригинальные правила интерпретации. Работа алгоритма построена на учете сугубо формальных свойств символов.

Во-вторых, для нас очевидна недостаточность простой «нарезки» текста по пробелам и знакам препинания, при которой точка, вопросительный и восклицательный знаки (так называемые терминальные знаки препинания) сигнализируют конец предложения, а пробел, запятая и др. — конец токена. При таком механическом подходе значительная часть информации может быть уте-

¹ См. также более раннюю публикацию [Клементьева, Скребцова, Суворов 2013].

ряна в силу неоднозначности знаков препинания, их способности выполнять разные функции (ср. [Безвербный 2000; Ровинская 2000; Кобзарева 2005]). Так, пробел и точка могут встречаться внутри чисел, больших тысячи, ср. 24 756 (= 24.756 = 24756) или 31 245 678 (= 31.245.678 = 31245678). В Великобритании и США, кстати, для этой цели обычно используется запятая, а в России запятая служит для отделения дробной части от целой в десятичных дробях (ср. 1,3 кг., 2,5 м.)². Помимо этого, точка используется в буквенных сокращениях (мер, весов и др.), которые могут встречаться в любом месте предложения, за исключением инициальной позиции. Точка и двоеточие применяются в обозначениях точного времени (7.15 или 7:15). Точка также встречается в обозначениях дат (24.12.12). Приведенные примеры, конечно, не исчерпывают все случаи, в которых знаки препинания не являются разделительным сигналом, но их достаточно, чтобы понять сложность задачи и неэффективность тривиальных решений.

Следует заметить, что формальный конец предложения не всегда знаменует законченность синтаксической структуры. В конструкциях с парцелляцией терминальные знаки препинания, по сути дела, выполняют функцию запятой, ср.:

Я требую амнистии. Я требую, чтобы она была полной и всесторонней. Без оговорок. Без ограничений (В. Гюго. Речь об амнистии в Сенате).

Ребятам так и сказали: хотите, мол, нажраться, поезжайте. Там все есть. И хлеб там есть. И картошка. И даже фрукты, о существовании которых наши шакалы и не подозревают (А. Приставкин. Ночевала тучка золотая).

Люда Голливуд мелькнула в кабинет начальника. Неслыханное дело: Чинков! Просит приема! У Фурдецкого! (О. Куваев. Территория).

— Я? Вам? Дал телефон? Что за ерунда! — не понимая, сказал Никитин. — Никогда! Я вам не давал никакого телефона! (Ю. Бондарев. Берег)³.

² Все примеры здесь и далее приводятся в оригинальной орфографии.

³ Это взятые из учебника, хрестоматийные примеры. Между тем парцелляция встречается не столь редко, как это может показаться на первый взгляд, и не только в художественной литературе. Следующий фрагмент заимствован из учебного пособия по стилистике русского языка (!), где он, как видно из текста, не используется в качестве иллюстрации, ср.: «В задачу этой книги не вхо-

При автоматической обработке текста соответствующие сегменты, несмотря на их пунктуационную самостоятельность, должны анализироваться вместе, как части одного предложения — только в этом случае они смогут получить адекватную интерпретацию. Сигналом к объединению членов парцеллированной конструкции служит выявляемая в ходе синтаксического анализа неполнота соответствующих парцеллятов. Их присоединение к базовой части конструкции происходит на этапе коммуникативного синтаксиса.

Графематический анализатор опирается на словарь символов, который создан авторами на основе ASCII-кодировки, т.е. включает буквы (русского и латинского алфавита, заглавные и строчные), знаки препинания, пробел, арабские цифры, дефис и некоторые другие часто используемые знаки компьютерной клавиатуры (знаки процента, доллара США, номера и т.д.). Использование символов ASCII-кода связано с тем, что разрабатываемый лингвистический процессор ориентирован на анализ текстов на русском языке, для которых в подавляющем большинстве случаев этой кодировки оказывается достаточно. В принципе, алгоритм может работать и с Unicode, правда, это потребует переделки словаря символов.

Важным преимуществом разработанного графематического анализатора является то, что он привязан к внешнему виду символа, а не к числовым кодам в компьютере. Это сохраняет возможность интерпретации при небрежном наборе текста или случайной опечатке (например, если вместо русского согласного *с* была напечатана английская буква *c*), а также в тех случаях, когда текст был отсканирован и распознан с ошибками.

Рассмотрим сказанное на примерах. Допустим, имеется последовательность из трех символов *САТ*. Каждый из них получит из словаря символов следующие варианты интерпретации⁴:

дит анализ разговорной речи, мы остановимся на трех функциональных стилях речи “книжной” — в первой части, и речи публичной, ораторской — во второй. На том, что в их современном состоянии есть “хорошего” или, наоборот, “плохого”. Как они пришли к их современному состоянию, как развивались, как “становились на ноги” и куда “идут” теперь. Какими языковыми средствами мы можем пользоваться, а каких лучше избегать, чтобы не выглядеть смешным в глазах читателя или слушателя — обо всем этом мы поговорим на страницах нашей книги» [Дунаевская 2010: 8].

⁴ Точнее, соответствующие значения по графическим признакам «тип символа», «алфавит», «регистр», «тип буквы».

С — латинский заглавный согласный⁵, русский заглавный согласный, обозначение температуры по Цельсию, римская цифра;

А — латинский заглавный гласный, русский заглавный гласный;

Т — латинский заглавный согласный, русский заглавный согласный.

Работа алгоритма основана на соединении смежных символов одного типа, поэтому на выходе будет получено два варианта: последовательность из русских букв САТ и последовательность из латинских букв САТ, — а интерпретации, связанные с обозначениями температуры и римской цифры, будут отброшены.

Другой пример: слово *сахарный*. При последовательном анализе символов слева направо неоднозначность, связанная с одинаковым написанием русских и латинских букв, сохраняется вплоть до шестой позиции. При обработке символа *н* обнаруживается, что в латинском алфавите отсутствует буква с подобным начертанием, следовательно, данная цепочка символов распознается как принадлежащая русскому алфавиту, и графематическая неоднозначность снимается.

Фундаментальный принцип работы графематического анализатора (и описываемого лингвистического процессора в целом) состоит в поэтапном объединении знаков во все более крупные сегменты. Сначала происходит соединение смежных символов одного типа в однородные последовательности, которые мы называем атомами (таким образом, рассмотренные выше последовательности из русских букв САТ и *сахарный*, а также последовательность из латинских букв САТ являются атомами). Формирование однородных последовательностей базируется на учете таких признаков символов, как «алфавит» (значения ‘кириллица’ / ‘латиница’), «регистр» (значения ‘заглавный’ / ‘строчный’) и др. Соответствующую информацию анализатор получает из словаря символов, так что, по сути дела, работа идет не с символом как таковым, а со значениями его признаков. Подчеркнем снова, что это важная особенность лингвистического процессора, характеризующая процедуру обработки текста на разных этапах.

Буквенные атомы всегда однородны с точки зрения алфавита. Они могут представлять собой последовательность строчных букв, последовательность заглавных букв, последовательность строчных

⁵ Более корректно: латинская заглавная буква, обозначающая согласный звук. См. также далее.

букв, открываемую заглавной буквой. При этом последовательность заглавных букв может означать либо аббревиатуру, либо шрифтовое выделение. Последовательность из начальной заглавной буквы и последующих строчных — это либо имя собственное, либо первое слово в предложении. Аббревиатуры, включающие строчные буквы (СПбГУ, ЦПКиО), представляются в виде нескольких атомов (в данных примерах — трех); их объединение происходит при следующих «проходах» графематического анализатора. Аналогично обстоит дело с графическим выделением отдельных букв в слове (ср. *рассы́пать* vs. *рассыпа́ть*) и использованием апострофа для обозначения ударного гласного (*рассы́пать*). Дефис обрабатывается как отдельный атом, поэтому слова типа *кое-где*, *кто-то* и т. п. на начальном этапе также предстают в «разорванном» виде, как и сокращения типа *г-жа*, *чл.-корр.*, *д-р*.

Другие примеры атомов включают, например, последовательность арабских цифр, последовательность пробелов, несколько одинаковых знаков препинания подряд (многоточие и др.). Все смешанные цепочки символов выступают в виде последовательности однородных атомов, ср.: $[24] \text{ } [oe]$; $[5] \text{ } [kz]$; $[ВУЗ] \text{ } [y]$; $[И]. [И]. [Иванов]$; $[Он] \text{ } [сказал]: [«] [Пошли] \text{ } [домой] \text{ } [»]$. Как видно из примеров, атом может быть образован одним-единственным символом. Такая высокая степень детализации анализа необходима для дальнейшей правильной интерпретации последовательностей атомов.

Подобно тому, как символам в соответствующем словаре приписаны значения признаков, атомы также получают свои значения признаков на основе правил интерпретации. Например, атом ВУЗ будет охарактеризован как последовательность из русских заглавных букв, гласных и согласных вперемешку, с указанием типов символов, ограничивающих его с обеих сторон (это могут быть знаки препинания, пробелы, дефис; справа к данному атому может также примыкать строчная буква кириллицы). Также будет указано место данного атома в предложении (начальное / среднее / конечное).

Затем из атомов формируются токены. В общем случае токен — это сегмент, который выступает в качестве цельного блока (единицы) на дальнейших этапах анализа. Сборка токена нередко осуществляется в несколько «проходов», например в случае комплексов из фамилии с инициалами (*А. С. Пушкин*, *А. Конан-Дойль*), аббревиатур из заглавных и строчных букв (СПбГУ), графических выделений (*звонит*), смешанных комплексов из цифр и знаков препинания (дат, обозначений точного времени, номеров телефонов и др.), циф-

ровых цепочек с буквенными сокращениями (70 км., 20 млн. долл.) или окончаниями (12-ти, 5-му), слов с дефисом и др. Токен может быть равен атому — ср. имена нарицательные, не имеющие в своем составе дефиса, или аббревиатуры, состоящие исключительно из заглавных букв.

Сборка токена осуществляется в соответствии с правилами интерпретации, предполагающими, в частности, его сравнение с набором шаблонов, таких как Ф.И.О. (с различными вариантами наполнения и расположения фамилии и инициалов, а также возможными сокращениями типа *г-н*, *тов.* в препозиции), аббревиатуры смешанного типа (состоящие из заглавных и строчных букв), пишущиеся через дефис слова, даты, цифровые последовательности (25,2; 46 792), цифровые последовательности с буквенными сокращениями (45 мин.) или окончаниями (5-ая, 5-я, 5-ти, 125я, 125ти), простые арифметические выражения ($2+2$, $450:15 = 30$) и некоторые другие. В итоге токену приписывается значение графического признака «тип токена» (например, ‘словоформа’, ‘аббревиатура из заглавных букв кириллицы’, ‘дата’ и пр.), а также значения таких признаков, как «место в предложении» (‘начальное’, ‘среднее’ или ‘конечное’), «тип ограничивающего слева символа» и «тип ограничивающего справа символа».

Покажем на некоторых примерах, как определяется тип токена. Так, состоящий исключительно из заглавных согласных токен (он же атом) СССР по признаку «тип токена» получит значение ‘аббревиатура из заглавных букв латиницы или кириллицы’; дальнейший поиск в словаре аббревиатур осуществляется на этапе атрибуции токенов — см. ниже. А вот цепочка символов МСХ, помимо аббревиатуры латиницей или кириллицей, может обозначать записанное римскими цифрами число 1110. Графематический анализ не имеет никаких данных в пользу какого-либо из этих трех вариантов, и снятие данной неоднозначности будет происходить позднее — в ходе синтаксического анализа.

Токен ВУЗ, включающий как гласные, так и согласные, в соответствии с правилами интерпретации может оказаться как аббревиатурой, так и шрифтовым выделением. Но если в постпозиции к цепочке ВУЗ примыкает строчная буква, то она однозначно распознается как падежная форма аббревиатуры (ср. ВУЗы, ВУЗе и др.). Таким токенам, как СПбГУ или УрГПУ, приписывается значение ‘аббревиатура’ за счет вкрапления строчной буквы в последовательность прописных. Цифры, разделенные знаками препинания, автоматически

объединяются в один токен, который затем сопоставляется с шаблонами дат и точного времени, фиксирующими наиболее частотные варианты написания.

Токены типа *где-то* идентифицируются в качестве словоформ, так как атомы слева и справа от дефиса представляют собой последовательности из русских гласных и согласных вперемешку. В отличие от них, символьные последовательности *г-жа*, *г-н*, *д-ру* и т.п. характеризуются как ‘буквенное сокращение’. Вообще, присоединение буквенных сокращений к соседним атомам с целью образования токена может осуществляться двумя способами: к предыдущему или последующему атому. Согласно правилам интерпретации, если буквенное сокращение стоит после цепочки цифр, но перед последовательностью строчных букв (‘некапитализированной словоформой’), оно присоединяется к цифрам (ср. *20 кг яблок*). Если же перед сокращением нет числа, а за ним идет последовательность строчных букв, открываемая заглавной буквой (‘капитализированная словоформа’), то сокращение относится к ней, ср. *г. Москва, тов. Иванов*. Данные правила охватывают наиболее частые случаи, однако не все последовательности атомов могут быть однозначно интерпретированы в рамках графематического анализа. Например, в цепочке *2012 г. Москва* буквенное сокращение *г.* может обозначать год и относиться к числу *2012*, а может обозначать город и примыкать к слову *Москва*.

Следует подчеркнуть, что приписывание токenu его типа носит предварительный характер. Это гипотеза, которая проверяется на следующем этапе анализа, а именно при атрибутировании токенов. Именно там, в частности, происходит поиск буквенных цепочек в словарях (словаре общих имен, словаре имен собственных, словарях сокращений и аббревиатур). Так, токен *Токарев*, стоящий в середине предложения, с самого начала будет правильно охарактеризован как имя собственное, так что проверка этой гипотезы даст положительный результат. Но если тот же токен расположен в инициальной позиции, первоначально ему будет приписан тип ‘словоформа’, и только процедура атрибуции позволит по итогам поиска в соответствующих словарях исправить его на ‘фамилия’.

Проверке подвергаются также и цифровые цепочки, и смешанные последовательности. К примеру, токены типа *24.12.2012* предположительно являются датами, но, чтобы подтвердить эту гипотезу, необходимо проверить составляющие его атомы с точки зрения соответствующих ограничений. Аналогичной процедуре подвергаются обозначения точного времени. Последовательностям заглавных

латинских букв (типа XVIII) при формировании токенов приписывается тип 'число', но далее осуществляется проверка их правильности с точки зрения допустимых символов и их сочетаний.

Как уже говорилось, помимо типа, для каждого токена, как и для атома, заполняются значения признаков, связанных с местом в предложении и типом ограничивающих символов. В отличие от атомов, токены не могут быть ограничены ни буквами, ни дефисами, ни цифрами — только знаками препинания и пробелами.

На следующем этапе работы графематического анализатора происходит сборка групп токенов, ограниченных знаками препинания. При этом следующие за знаком препинания пробелы удаляются, поскольку они являются сугубо конвенциональными и не несут функциональной нагрузки. Собранные таким образом группы токенов называются «частями предложения» (графическими, а не синтаксическими). Часть предложения может совпадать с одиночным токеном, ср. одиночное вводное слово или член однородного ряда и др.

В соответствии с общим принципом работы анализатора части предложения получают описания в терминах занимаемого места и ограничивающих символов. Так, пушкинское предложение *Мороз и солнце; день чудесный!* состоит из двух частей: первая стоит в начале предложения, ограничена слева началом текста и точкой с запятой справа, а вторая стоит в конце предложения и (после удаления конвенционального пробела) ограничена точкой с запятой слева и восклицательным знаком справа. Часть предложения также может быть охарактеризована с точки зрения ее типа: например, 'скобочная структура', 'прямая речь' и др.

Наконец, части предложения объединяются в предложение. В общем случае эта процедура основана на поиске терминального символа (точки, вопросительного или восклицательного знака), но следует упомянуть о двух исключениях из этого правила. Во-первых, как указывалось выше, точка может быть составной частью цифровых шаблонов и буквенных сокращений, следовательно, наличие точки не означает автоматически конец предложения. Во-вторых, терминальные знаки препинания нередко встречаются внутри прямой речи, при этом бывает целесообразно рассматривать в качестве предложения всю структуру, включающую как авторскую речь, так и речь персонажа, ср.:

«Да проститься же надо было!..» — понял он, когда крытая машина взбиралась уже на взвоз (В. Шукшин. Осенью).

«Вдруг посеешь, — думал Семен, — а вырастет обыкновенный ячмень. Скорее всего, так и получится. Но попробовать надо. И, главное, поискать еще. Неужели в целом поле он был один?» (В. Солоухин. Счастливы колос);

— Превосходно, Уотсон, — воскликнул Холмс, — мне кажется, говорю вам это искренне, что вы недалеко от истины. Вы видите сами, все карты у нас в руках, и теперь нам надо спешить, пока не случилось непоправимое. Если время позволит, мы захватим их непременно (А. Конан-Дойль. Случай с переводчиком).

В подобных случаях отдельные графические предложения внутри прямой речи рассматриваются нами в качестве специальной единицы, которую мы называли подпредложением. И подпредложениям, и предложениям приписываются значения, связанные с их позицией в более крупной единице (соответственно предложению и абзаце) и типами ограничивающих символов.

Самым крупным сегментом графематического анализа является абзац, который, однако, иногда может быть равен предложению или даже еще меньшим сегментам. Расстановка символа «конец абзаца» происходит в рамках первичного композиционного анализа (см. выше). Однако следует заметить, что не всегда знак конца абзаца означает смысловую законченность: достаточно указать на списочные структуры после двоеточия, где каждый новый пункт (нумерованный или определенным образом маркированный) располагается с новой строки и начинается со строчной буквы. Фактически мы имеем здесь одно предложение, графически разбитое на несколько абзацев. (Можно усмотреть некоторую аналогию с членами парцелированной конструкции, отделенными друг от друга терминальными знаками препинания.)

* * *

Основным источником проблем, возникающих при графематическом анализе текста, является неоднозначность знаков препинания, зависимость их функции от контекста. В этой связи следует прежде всего упомянуть кавычки, которые могут использоваться в разных целях, а именно: для выделения цитат, отдельных слов (устарелых, непривычных, употребляемых иронически, терминов и пр.), условных наименований (подробнее о семантике кавычек см. [Зализняк 2007]). Наши правила интерпретации охватывают следующие ситуации:

- Если в кавычки заключен одиночный токен, по признаку «тип» ему приписывается значение 'название' (если он написан с заглавной буквы и не расположен в начале предложения, ср. *издательство «Наука», союз «Тартар»*) или 'термин'⁶ (если он написан со строчной буквы и не расположен в начале предложения, ср.: *Так и появилась болезнь под названием «арахнофобия»*). Написанный с заглавной буквы одиночный токен в кавычках в инициальной позиции может быть как тем, так и другим;
- Если новый абзац открывается группой токенов в кавычках (первый токен с заглавной буквы), за которой следует запятая или тире, а также если подобная группа токенов в кавычках расположена после двоеточия, она представляет собой прямую речь;
- если группа токенов в кавычках не занимает начальной позиции в предложении, не предваряется двоеточием и начинается со строчной буквы, она атрибутируется в качестве цитаты. Например: *Говоря о поэзии Пушкина, Н. А. Добролюбов писал, что «в его стихах впервые сказалась нам живая русская речь, впервые открылся нам действительный русский мир».*

За пределами данных правил оказываются многословные наименования, например названия литературных произведений: *«Кому на Руси жить хорошо?»*, *«Повесть о настоящем человеке»*, фильмов: *«Москва слезам не верит»*, *«Ирония судьбы, или С легким паром»*, телевизионных игр: *«Кто хочет стать миллионером?»*, *«Что? Где? Когда?»*, газет и журналов: *«Аргументы и факты»*, *«Русский язык за рубежом»*, *«Знание — сила»*, а также лозунги: *«Вся власть советам!»*, *«Слава КПСС!»* и др. Подобным группам токенов графематический анализатор не может приписать тип; их атрибутирование происходит при синтаксическом анализе.

Графематический анализ предложений, содержащих кавычки, может быть дополнительно осложнен тем обстоятельством, что кавычки, как известно, бывают разного рисунка. В том числе есть и такая, в которой открывающие кавычки по внешнему виду совпадают с закрывающими. В ситуации, которую условно можно обозначить следующим образом: " " " ", — возможны следующие две интерпретации:

⁶ Данное обозначение следует понимать в расширенном смысле, а именно: термин или слово, к появлению которого в тексте автор в силу тех или иных причин хочет привлечь внимание (это может быть окказионализм, экзотизм, устаревшее слово, шутовское или ироническое наименование и т. д.).

- вложенные одна в другую структуры, ср.: Пушкин писал Дельви-гу: "Жду "Цыганов" и тотчас тисну";
- два фрагмента, следующие друг за другом, ср. Читали ли вы статью в "Известиях" "Куда мы идем?".

Автоматически распознать эти варианты в рамках графематического анализа невозможно.

Трудный случай представляет собой также использование в тексте непарных кавычек, ср. ЗАО «Издательский дом «Комсомольская правда», ОАО «НПП «Буревестник», работа В. И. Ленина «О карикатуре на марксизм и об "империалистическом экономизме"».

Многозначность свойственна и другим знакам препинания. Особо высок индекс неоднозначности у дефиса, который теоретически может трактоваться и как буква, и как знак препинания [Крылов, Старостин 2003: 354]. Кроме случаев его использования в составе слова и в качестве знака переноса, дефис может быть сигналом сокращенного написания слов (как простых, так и сложных, ср. *г-н, д-р, 150-летие, 5-го*), применяться при записи конструкций с однородными членами, имеющими общий конечный элемент (например, *теле- и видеофильмы; не двух-, а трехэтажный дом; как водо-, так и газоснабжение*), употребляться в выразительных целях, быть соединительным символом для составных существительных (типа *девочки-подростки*), числовых шаблонов (таких как даты, номера телефонов) и др. К тому же в компьютерной записи дефис (или два дефиса подряд) иногда применяется для обозначения тире. Графематический анализатор должен уметь распознавать все это многочисленные случаи и действовать соответственно: приводить написание слова к полной форме, строить слово по аналогии с соответствующим однородным членом, объединять последовательность знаков в единый смысловой блок и т. д.

Алгоритм стремится к максимально полному учету контекста любого символа, с тем чтобы снимать неоднозначность или, по крайней мере, ее сокращать. Это достигается за счет многоступенчатой процедуры анализа контекста на основе разработанных правил интерпретации. Помимо сокращения числа ошибочных интерпретаций, данная процедура позволяет осуществлять глубокий структурный анализ текста, в том числе распознавание вложенных структур, которые могут быть оформлены такими парными знаками препинания, как кавычки, скобки, тире и даже дефисы. В целом можно сказать, что мы стремимся «выжать максимум» из сугубо формального анализа на графематическом уровне.

Глава 2

АТРИБУЦИЯ ТОКЕНОВ

МОРФОЛОГИЧЕСКИЙ АНАЛИЗ БУКВЕННЫХ ТОКЕНОВ

Цель данного этапа работы процессора заключается в том, чтобы проверить сделанные на этапе графематического анализа гипотезы о типе токенов, а также приписать им значения релевантных морфологических признаков. Если токеном является словоформа русского языка, речь идет о процедуре, которую принято называть морфологическим анализом. Но, поскольку в текстах встречаются и другие виды токенов (см. выше), мы пользуемся более общим термином «атрибуция», подчеркивающим приоритетную задачу данного этапа — определение типа для всех токенов¹. Решение данной задачи является необходимой предпосылкой успешного синтаксического анализа.

Рассмотрим процедуру атрибуции для токенов различной природы. Наиболее частотными в естественно-языковых текстах, несомненно, являются буквенные токены. Если не принимать во внимание регистр, буквенный токен может быть самостоятельной словоформой, аббревиатурой или частью устойчивого словосочетания. Соответственно, у нас имеются три различных словаря, в которых можно искать текущий токен. Заметим, что аббревиатуры также могут состоять из нескольких частей (ср. *МВД РФ*), поэтому есть смысл рассматривать подобные составные аббревиатуры также в качестве устойчивых сочетаний, игнорируя регистр. Тогда для всех буквенных токенов процедура атрибуции начинается одинаково: с обращения к словарю устойчивых неизменяемых словосочетаний².

Под устойчивым словосочетанием мы понимаем фиксированную последовательность словоформ или аббревиатур, выражающую еди-

¹ В некоторых системах автоматической обработки языка выделяется специальный этап постморфологии, посвященный обработке имен собственных, аббревиатур, сокращений, цифровых последовательностей и др.

² Так, независимо от того, как в тексте выглядит текущий токен — *МГУ*, *Мгу* или *мгу*, — в словаре устойчивых словосочетаний все равно ищется некапитализированная буквенная цепочка *мгу*, с целью дальнейшего выяснения, не является ли данный токен начальным компонентом устойчивого словосочетания.

ное понятие и функционирующую в качестве цельного блока³. Под это формальное определение подпадают такие содержательно разные вещи, как неизменяемые фразеологические единицы, вводные обороты, сложные союзы, составные наречия, местоимения, предлоги и частицы, а также аббревиатуры из двух и более компонентов (например, *ГУ МВД РФ*). Естественно, данный список не включает составные словоформы, образующиеся по правилам (аналитические формы сравнительной и превосходной степени наречий и прилагательных, аналитические формы глаголов, составные числительные), в силу того что они представляют собой открытый список изменяемых форм. На случай вариативности структуры устойчивого словосочетания (например, *не иначе/не иначе как*) в словаре предусмотрены все варианты его реализации.

Итак, алгоритм проверяет, не является ли текущий буквенный токен частью устойчивого словосочетания. Если найдено устойчивое словосочетание, начинающееся с данного токена, далее осуществляется сравнение остальной части этого словосочетания с токенами, которые в тексте непосредственно следуют за данным. (Поскольку процедура атрибуции последовательно осуществляется слева направо, от токена в инициальной позиции вплоть до токена в финальной позиции, достаточно всякий раз анализировать только правый контекст.) Когда устойчивое сочетание содержит более чем две словоформы/аббревиатуры, сопоставление требует нескольких шагов. Если последовательность словоформ найдена в словаре, ей приписывается соответствующее значение по грамматическому признаку «часть речи». Например:

на босу ногу — нар
в том числе — союз

Если найденное устойчивое сочетание представляет собой последовательность аббревиатур, словарь сопоставляет ей следующую информацию: морфологические характеристики составной аббревиатуры, определенные на основе анализа ее функционирования в тексте, лемму опорного слова всего сочетания и позицию ударного гласного в данной лемме. Морфологические характеристики аббревиатуры могут не совпадать с соответствующими характеристиками опорного слова. При вариативности морфологических характеристик приводятся все варианты, ср.:

³ Ср. понятие безусловного оборота в работе [Апресян и др. 1992: 34].

гу мвд рф	сущ; неод; ср ⁴	управление	3
мвд рф	сущ; неод; мр	министерство	3
мвд рф	сущ; неод; ср	министерство	3

Если рассматриваемый токен совокупно со своим правым контекстом не найден в словаре устойчивых словосочетаний, подключается предварительная (сделанная на этапе графематического анализа) гипотеза о его типе. Одиночный буквенный токен, в зависимости от своего состава, может быть либо словоформой (ср. *Мгу, мгу*), либо аббревиатурой (*МГУ*). Рассмотрим тот случай, когда токен предположительно является словоформой (имени нарицательного или собственного).

Для таких токенов процедура атрибуции продолжается поиском словоформ в словаре, который, наряду с нарицательными существительными, содержит частотные имена, фамилии, отчества, географические названия, наименования организаций и пр. в полном наборе их словоизменительных вариантов. В основу словаря словоформ первоначально был положен находящийся в открытом доступе морфологический словарь проекта АОТ⁵. В дальнейшем, однако, он был дополнен материалами других словарей и существенно переработан как в лингвистической, так и в алгоритмической части. Можно сказать, что из проекта АОТ заимствована главным образом сама идеология компактного хранения и быстрого поиска.

В словаре словоформ регистр игнорируется, и все заголовочные слова пишутся строчными буквами. Это обусловлено тем обстоятельством, что регистр не всегда позволяет однозначно определить, является ли текущий токен именем собственным или нарицательным. Так, слово в инициальной позиции в предложении (или в прямой речи) пишется с заглавной буквы независимо от того, собственное это имя или нарицательное, ср.: *мга* ('туман, мгла') и *Мга* (река и населенный пункт в Ленинградской области). В подобных случаях словарь словоформ выдает оба варианта:

мга	1	сущ; лок ⁶ ; неод; жр, им; ед
мга	1	сущ; неод; жр, им; ед

⁴ Неодушевленное существительное среднего рода. Ниже, соответственно, неодушевленное существительное мужского рода.

⁵ URL: <http://www.aot.ru> (дата обращения: 04.04.2018).

⁶ Локатив (населенный пункт, река, гора и др.).

Примеры подобной неоднозначности не так уж редки, ср. *пашу́* — *Пашу*, *ми́лá* — *Мила*, *ве́ра* — *Вера*, *ю́ля́* (деепр. от *юлить*) — *Юля*, *ве́на* — *Вена* и т. д.

Итак, словарь сопоставляет каждой словоформе ее лемму, порядковый номер ударного слога, частеречный разряд и грамматические характеристики (отдельно по словоизменительным и классифицирующим морфологическим категориям), а именам собственным дополнительно приписывает условный маркер ('локатив', 'фамилия', 'имя' и др.). Несклоняемые географические названия получают характеристику 'неизм', ср.:

тбилиси 2 сущ; неизм; лок; неод; ед; мр

Их род определяется по роду нарицательного слова, для которого имя собственное служит наименованием (в нашем примере — по слову *город*). Для географических названий, оканчивающихся на *-о* и *-е* и не имеющих в словаре указания на тип населенного пункта, указывается средний род (ср. *украинское Ровно*, но *сибирский Кемерово*). Кроме того, для многих географических названий на *-о* (типа *Останкино*) предусмотрено два варианта — изменяемое и неизменяемое существительное.

Неизменяемые фамилии получают описание, соответствующее их способности относиться к лицу мужского или женского рода, ср.:

петренко 2 сущ; неизм; фам; неод; мр/жр

Для подавляющего большинства словоформ словарь выдает несколько вариантов морфологического разбора. Например, для известного примера *стекло* будут получены следующие три варианта атрибуции:

стекло	2	сущ; неод; ср	им; ед
стекло	2	сущ; неод; ср	им; вин
стечь	2	гл; нп ⁷ ; соверш	прош; ср

Множественность полученных результатов, разумеется, усложняет дальнейший (синтаксический) анализ. Иногда число вариантов удастся сократить, опираясь на значения графических признаков соответствующих токенов. Так, если токен занимает неинициальную позицию в предложении и написан с заглавной буквы, из списка вариантов выбирается тот, который является словоформой имени

⁷ Непереходный глагол.

собственного (в примере со словом *мга* — первая строчка). Если результаты графематического анализа говорят о том, что токен имеет в своем составе одну неначальную капитализированную гласную, имеет смысл сопоставить порядковый номер соответствующего слога в токене с заданной в словаре позицией ударения (ср. *хло́пок* или *хлопо́к*): совпадение номеров подтвердит мысль о том, что таким образом помечен ударный слог, и позволит убрать лишние варианты атрибуции. Аналогичная процедура применяется, когда для указания ударного гласного в буквенных токенах используется апостроф (ср. *хло́пок* или *хлопо́к*).

Таким образом, если первоначально (при обращении к словарю словоформ) графематическая информация о текущем буквенном токене игнорируется (поскольку в словаре все заголовочные слова записаны строчными буквами и апострофы внутри токена удалены), то позднее, когда мы уже имеем результаты атрибуции, эта информация начинает работать для исключения заведомо неверных вариантов.

Обратимся теперь к рассмотрению работы алгоритма в случае, когда не найденный в словаре устойчивых словосочетаний токен предположительно является аббревиатурой. Происходит обращение к специальному словарю аббревиатур и буквенных сокращений (о последних см. ниже). Заголовочные единицы в нем записываются с соблюдением регистра, так как в ряде случаев он значим, ср. сокращения *Вт.* (ватт) и *вт.* (вторник). Если имеются различные способы записи одной и той же аббревиатуры (сокращения), в словаре приводятся все варианты, ср. *ВУЗ* и *вуз*.

В словаре аббревиатурам приписаны их морфологические характеристики, «вычисленные» с опорой на корпус текстов. В случае вариативности рода приводятся все варианты. Словарь также сопоставляет каждой аббревиатуре ее опорное (стержневое) слово и позицию ударного гласного в нем. Последнее необходимо для того, чтобы на этапе синтаксического анализа (см. ниже) получить семантические характеристики для данного токена по соответствующему словарю. Морфологические характеристики аббревиатуры и опорного слова зачастую не совпадают, ср.:

НИИ	сущ; неод; мр	институт	3
НИИ	сущ; неод; ср	институт	3
МИД	сущ; неод; мр	министерство	3
БАМ	сущ; неод; мр	магистраль	3

В некоторых случаях приписывание опорного слова осуществляется искусственно. Это касается, в частности, заимствованных из других языков звуковых аббревиатур: так, для *НАТО*, *ЮНЕСКО*, *ФИФА* опорным словом будет *организация*, а для *Би-Би-Си* — *радиостанция*. Для таких аббревиатур, как *СНиП* (строительные нормы и правила) делается выбор в пользу одного из членов однородного ряда.

Если искомого токена в словаре аббревиатур нет, предварительная гипотеза отменяется и производится его поиск в словаре словоформ. Ведь капитализация возможна не только в случае аббревиатур, но и текстовых выделений, а также в заголовках.

Различного рода слоговые аббревиатуры (они же сложносокращенные слова) типа *Минкульт*, *партком*, *Роспотребнадзор*, *телеинтервью*, *запчасть*, *завкафедрой* обрабатываются как обычные словоформы. Сначала происходит обращение к словарю словоформ (наиболее частотные из них туда занесены). Если сложносокращенное слово там не найдено, к работе подключается механизм морфологического предсказания (см. ниже).

Особо следует сказать о таких токенах, как *VIP-персона*, *SMS-сообщение*, *DVD-диск*: первая часть, оформленная латиницей, отбрасывается, и атрибуция токена производится по второй части, которая ищется в словаре словоформ.

АЛГОРИТМ МОРФОЛОГИЧЕСКОГО ПРЕДСКАЗАНИЯ

Рассмотрим теперь тот случай, когда буквенный токен не обнаружен ни в одном из трех упомянутых словарей, т.е. является так называемой несловарной словоформой [Ляшевская 2007; Клышинский 2007; Черненьков 2010]. В литературе описаны различные подходы к их морфологическому анализу (см., например, [Ермаков, Плешко 2004; Сокирко 2010]).

В нашем лингвистическом процессоре применяются следующие эвристические приемы определения грамматических характеристик неопознанных буквенных токенов. Если токен содержит дефис, прежде всего делается попытка удалить дефис (а вдруг это знак переноса?) и искать соответствующую цепочку символов без дефиса в словаре словоформ.

Если поиск не дал результата, алгоритм ищет части токена (до и после дефиса) по отдельности. Здесь возможны разные ситуации.

Если найдена только вторая часть, морфологическая атрибуция всего сложносоставного слова⁸ производится по ней, ср. *итало-австрийский, физико-технический, социал-демократ*. Так же обстоит дело, если в словаре содержатся обе части сложносоставного слова, причем вторая часть является прилагательным (например, *светло-голубой, один-единственный*). Проблем не возникает и в случае такой редупликации, при которой компоненты полностью совпадают, ср. *еле-еле, чуть-чуть, белый-белый*.

Трудности появляются, когда в словаре мы находим только первую часть сложносоставного слова, ср. *рад-радехонек, старый-престарый, мало-помалу*. Казалось бы, можно осуществлять атрибуцию по этой части, однако в общем случае это неверно: например, в словах типа *белым-бело* это приведет к тому, что вместо неизменяемого наречия данное слово будет атрибутировано как прилагательное единственного числа, мужского или среднего рода в творительном падеже. Как поступать в такой ситуации, пока неясно.

Нами была подробно рассмотрена более частная проблема — определение морфологических характеристик сложносоставного существительного, обе части которого найдены в словаре⁹. Понятно, что когда оба компонента относятся к одному грамматическому роду (*вагон-ресторан, бал-маскарад, генерал-майор, девочка-украинка, старуха-процентщица*), род сложносоставного существительного будет таким же. Что касается словоизменительных характеристик, то их разумно определять по форме второго компонента, поскольку первый может быть неизменяемым или не склоняться в составе данного сложносоставного слова (например, *кофе-машины, генерал-майора, но вагона-ресторана, дому-музею*).

⁸ Под сложносоставными словами мы, вслед за орфографическим справочником (Соловьев 1997: 740–741), понимаем сложные слова, пишущиеся через дефис; соответственно, собственно сложные слова — это те, которые имеют слитное написание. Иная точка зрения представлена в (Русская грамматика 1980: т. 1: 253), где разграничение между сложными и сложносоставными словами проводится исходя из особенностей их склонения: сложные слова содержат несклоняемый первый компонент, в то время как сложносоставные представляют собой словосочетания, в которых склоняются обе части. В то же время сами авторы цитируемого источника признают, что граница между этими двумя типами слов не вполне жесткая (Русская грамматика 1980: т. 1: 253). В связи с этим опора на материальный признак представляется нам более обоснованной и, вдобавок к тому, более удобной для изложения, поскольку мы рассматриваем здесь исключительно слова с дефисным написанием.

⁹ См. [Скребцова 2014].

Задача перестает быть тривиальной, когда компоненты сложно-составного существительного имеют разный грамматический род. Род определяет согласование существительного с определением и сказуемым (когда такое существительное выступает в роли подлежащего), а также выбор анафорического местоимения (*он, она, оно, который (-ая, -ое)* и др.). Очевидно, что определение этой грамматической характеристики важно для последующих этапов работы лингвистического процессора, связанных с анализом предложения и связного текста. Разумеется, можно приписывать сложносоставным существительным типа *диван-кровать, шкаф-купе, бизнес-сообщество* по два грамматических значения рода, но это повлечет за собой существенный рост синтаксической неоднозначности.

Проблема определения рода сложносоставных существительных не получила широкого освещения в литературе. Авторы теоретических статей обычно апеллируют к понятию приложения, что в контексте автоматического анализа неуместно. В статье [Большаков, Большакова 2012] предложен примитивный практический подход, согласно которому определять род и число сложносоставного существительного нужно исходя из соответствующих характеристик первого компонента. Его несостоятельность легко показать на многочисленных примерах.

В пособии Д. И. Розенталя сделан упор на понятие ведущего слова. Автор, в частности, пишет: «...ведущим (определяющим) является то слово, которое выражает более широкое понятие или конкретно обозначает предмет. <...> Как показывают примеры, обычно в этих случаях на первом месте стоит ведущее слово, с которым и согласуется сказуемое. Если же на первом месте стоит не ведущее слово, то при изменении падежной формы подобных сочетаний первая часть не изменяется» [Розенталь 1998: 239]. В качестве иллюстраций соответственно приводятся существительные *девушка-агроном* vs. *плащ-палатка*: ср. *девушки-агронома*, но *плащ-палатки*.

В целом соглашаясь с приведенными суждениями, заметим, однако, что даже человеку (а тем более компьютеру) бывает нелегко определить ведущее слово. Так, в существительных *хлеб-соль, братья-сестры, руки-ноги* компоненты семантически равноправны (объемы соответствующих понятий не пересекаются) и в одинаковой степени конкретны. В существительных типа *бизнес-сообщество, рок-группа, интернет-телевидение*, построенных по заимствованной из английского языка модели, тоже не наблюдается отмеченных смысловых отношений между компонентами. В самом деле, какое

понятие более широкое — *рок* (как музыкальное направление) или *группа* (как коллектив)? Какое слово обозначает конкретный предмет — *Интернет* или *телевидение*? К тому же сам автор цитированного пособия признает отдельные «колебания в согласовании глагола-сказуемого со сложносоставными существительными, в которых одна часть по функции напоминает приложение, ср. *кафе-столовая открыто/открыта*» [Розенталь 1998: 238–239].

Довольно замысловатые правила определения рода сложносоставного слова содержатся в работе [Волынец 2003: 11]:

- 1) если составное существительное обозначает лицо, то его род определяется по слову, указывающему на пол лица — *женщина-скульптор создала...*, *чудо-богатырь поскакал...*;
- 2) если составное существительное обозначает неодушевленные предметы, то род определяется по роду первого слова — *музей-квартира открыт...*, *школа-интернат построена...*, *кресло-кровать стояло в углу...*;
- 3) если в сложносоставном слове одно существительное несклоняемое, то род определяется по роду склоняемого слова — *пресс-конференция прошла успешно*, *кафе-столовая открыта...*

Однако использование этих правил в автоматическом морфоанализе наталкивается на препятствия. Во-первых, информация о том, обозначает ли то или иное существительное лицо, указывает ли оно на пол лица, представляет собой семантические характеристики, которые, как правило, приписываются словоформам на более поздних этапах анализа. В нашем словаре словоформы характеризуются только признаком морфологической одушевленности/неодушевленности, которого в данном случае недостаточно. Таким образом, пункт 1) не несет полезной для нас информации.

Во-вторых, в процитированном выше методическом пособии ничего не говорится об обозначениях живых существ типа *рыба-меч*, *бабочка-адмирал* — похоже, что у них грамматический род определяется в соответствии с пунктом 2), т. е. так же, как у существительных, обозначающих неодушевленные предметы. Получается, что указанный выше признак морфологической одушевленности/неодушевленности, которым характеризуются словоформы в нашем словаре, опять нерелевантен.

Наконец, пункт 3) при ближайшем рассмотрении оказывается неточным (чтобы не сказать — неверным), так как есть много

слов, у которых первая часть сама по себе в отдельности склоняется, но в сложносоставном существительном становится неизменяемой, ср. *бизнес-школа, рок-группа, плащ-палатка, горе-чиновник, чудо-сад*. Формально они относятся к пункту 2), и их грамматический род должен определяться по роду первого компонента, но в действительности дело обстоит ровно наоборот: род сложносоставного слова совпадает с родом второго компонента. Значит, по сути, следовало бы причислять их к третьему типу. Но тогда возникает задача формального определения класса существительных, которые утрачивают способность склоняться, когда выступают в качестве первого компонента сложносоставных существительных. Непонятно, как она может быть решена¹⁰.

Рассуждая о понятии морфосинтаксического локуса — элементе словосочетания, в котором морфологически выражается его связь с более широким контекстом, — Я. Г. Тестелец констатирует, что синтаксической вершиной в аппозитивных конструкциях может быть как первый элемент сочетания (*Мы увидели славный город-герой*), так и второй (*В путь-дорогу дальнюю я тебя отправляю*); более того, оба существительных могут оказаться морфосинтаксическим локусом для разных грамматических явлений, ср. *Черная птица-шофер на лету отвинтил правое переднее колесо* (М. А. Булгаков. *Мастер и Маргарита*): определение согласуется с одним существительным, а сказуемое — с другим [Тестелец 2001: 84–85].

Подводя итог затянувшимся размышлениям, приходится констатировать, что в настоящий момент мы не видим способа в полном объеме решить задачу вычисления грамматического рода (а следовательно, и словоизменительных характеристик) для всех сложносоставных существительных. В ряде случаев приходится допускать возможность альтернативного рода, что в итоге ведет к увеличению общего числа вариантов морфоанализа.

¹⁰ Любопытно, что ситуация, в которой компоненты сложносоставного существительного имеют разное грамматическое значение числа, разрешается гораздо проще. Если эта разница происходит из-за того, что один из компонентов представляет собой существительное *pluralia tantum*, род составного слова (и его словоизменительные характеристики) определяются по другому компоненту, ср. *юбка-брюки, ясли-сад, часы-будильник*. Различие в числе также возможно тогда, когда первый компонент сложносоставного слова не склоняется, а второй стоит в форме множественного числа, ср. *чудо-йогурты, пресс-службы, генерал-майоры* — тогда, очевидно, род и другие морфологические характеристики сложносоставного существительного вычисляются по второму (изменяемому) компоненту.

Схожие проблемы возникают при попытке определения одушевленности/неодушевленности некоторых сложносоставных существительных, ср. *город-герой*, *страна-поставщик*, *бизнес-партнер* и т.д. Заметим, что одна и та же модель построения (неодушевленное существительное + одушевленное существительное) дает в приведенных примерах разный результат: первые два сложносоставных существительных являются неодушевленными, а последнее — одушевленным. Как и в случае с морфологическим родом, правильное определение значения данной грамматической категории важно для дальнейших этапов автоматического анализа; между тем литературы по данному вопросу практически нет (подробнее см. [Скребцова 20156]).

Обратимся теперь к ситуации, когда в составе несловарной словоформы нет дефиса или, несмотря на описанные выше приемы, в словаре не удалось ничего найти. В этом случае к работе подключается алгоритм морфологического предсказания. Он опирается на списки псевдокорней (ср. *гран*, *границ*, *границ*, *гранк*, *гранул* и т.п.), «концовок» (ср. *-зала*, *-жегши*, *-обраны*, *-этому*, *-нувших*, *-атизм*, *-итель*, *-чивый* и т.п.), а также приставок (в том числе двойных) и префиксоидов, созданные нами на базе «Толково-словообразовательного словаря» (Ефремова 2000) и проекта АОТ. Разработанный алгоритм пытается, двигаясь от конца словоформы, последовательно отделять возможные «концовки» и осуществлять поиск остатка в списке псевдокорней; аналогично от начала словоформы он пытается отсечь потенциальные префиксы и также искать подходящий корень. Если эта процедура дает результат, для словоформы восстанавливается предположительная лемма и определяются вероятные грамматические характеристики.

Разумеется, далеко не для всех неопознанных словоформ данный процесс заканчивается успешно. Во-первых, попытки отделения аффиксов могут оказаться неэффективными, в том смысле, что ни с начала, ни с конца словоформы не будет найдено подходящих фрагментов. Возможно, что некоторые грамматические сведения о такой словоформе впоследствии даст синтаксический анализ. Например, в следующей последовательности из пяти слов:

[прил], [прил] [и] [неопозн] [сущ] —

неопознанная словоформа не может быть ничем иным, кроме как однородным определением к существительному. (Кстати говоря, синтаксический анализ позволяет «задним числом» не только

приписать морфологические характеристики неопознанным словоформам, но и уточнить их для уже атрибутированных токенов, например определить падеж для аббревиатур, число для сокращений, тип числительного для записанных цифрами чисел.) Хотя такие информативные контексты в целом редки, синтаксический анализатор в любом случае стремится прикрепить неопознанный токен к какой-нибудь конструкции.

Во-вторых, не так уж редки случаи ложной омонимии флексий (подробно рассмотренные в [Мищенко, Калугина 2002]), приводящие к неверному морфемному членению. С этой проблемой можно было бы справиться, разработав полноценный морфологический анализатор, осуществляющий морфемное членение словоформ, однако на сегодняшний день он не готов (подробнее см. ниже параграф «Атрибуция других типов токенов»).

АТРИБУЦИЯ ДРУГИХ ТИПОВ ТОКЕНОВ

Обратимся теперь к процедуре атрибуции для токенов, которые, наряду с буквами, включают цифры и/или различные знаки. Начнем с буквенных сокращений (типа *кг*, *км*, *др.*, *г.*, *г-жа*). Как уже указывалось, они входят в словарь аббревиатур и сокращений. Если встречаются разные способы графического оформления сокращения (*р.*, *руб.*; *млн.*, *млн.*; *ж.д.*, *ж.-д.*), в словаре приводятся все варианты. Для каждого сокращения приводится его «расшифровка» (не всегда совпадающая с леммой), порядковый номер ударного слога (для однословных «расшифровок»), а также (по возможности) частеречная характеристика и морфологические характеристики, ср.:

кг	килограмм	3	сущ; но; мр
др.	другие	2	прил; мн
пп.	пункты	1	сущ; но; мр; мн
т. е.	то есть		союз
т. п.	тому подобное		
ж.-д.	железнодорожный	5	прил
с.-х.	сельскохозяйственный	4	прил
ув.	уважаемый	3	прил
тов.	товарищ	2	сущ

Недостающая морфологическая информация может быть восполнена на этапе синтаксического анализа. Составные сокращения (типа *млн. долл.*, *тыс. км*), подобно составным аббревиатурам, вхо-

дят в словарь устойчивых словосочетаний. Аналогично мы поступаем с аббревиатурами типа *км/ч*, *кг/кв.м*.

Кроме того, для сокращений введен специальный признак, призванный охарактеризовать то, к чему синтаксически примыкает то или иное сокращение. Возможные значения данного признака следующие: ‘к числу’, ‘к имени’ (с частным вариантом ‘к имени собственному’) и ‘самостоятельное сокращение’ (оно же — ‘концевое’, например *т. д.*, *т. п.*, и др.).

Это помогает снимать или по крайней мере сокращать неоднозначность. Так, *г.* может обозначать год, грамм, город, гору и господин. Благодаря указанному признаку, удастся развести значения ‘год’ и ‘грамм’, с одной стороны, и ‘город’, ‘гора’, ‘господин’ — с другой. Далее, если перед сокращением стоит записанное арабскими цифрами четырехзначное число, скорее всего, оно обозначает год (для одно-, двух- и трехзначных арабских чисел, обозначающих год, после *г.* нередко идет еще одно сокращение — *н. э.* или *до н. э.*). Для цепочек типа *г. Москва*, если *Москва* найдена в словаре словоформ и опознана как локатив, остаются варианты интерпретации ‘город’ и ‘гора’. Что же касается, к примеру, цепочки *г. Пушкин*, то она может быть как географическим названием (ср. пригород Санкт-Петербурга с таким названием), так и обозначением лица.

Атрибуция подобных токенов (*2012 г.*, *г. Москва*, *г. Эльбрус*) предполагает сравнение с так называемыми шаблонами. Шаблоны — это цепочки символов смешанной природы, которые могут включать буквы, цифры и другие знаки. Они были выделены для обработки часто встречающихся последовательностей символов, таких как даты, обозначения точного времени, буквенные сокращения с именами собственными в постпозиции или цифрами в препозиции, аббревиатуры или последовательности цифр с буквенными окончаниями, интернет-адреса, обозначения веков латинскими цифрами, арифметические выражения и др.

К примеру, учитываются разнообразные способы записи дат:

дд.мм.гг; *дд.мм.гггг*; *дд/мм/гг*; *дд/мм/гггг*; *дд-мм-гг*; *дд-мм-гггг*;
мм.дд.гг; *мм.дд.гггг*; *мм/дд/гг*; *мм/дд/гггг*; *мм-дд-гг*; *мм-дд-гггг*
 (американские способы записи дат).

При этом алгоритм проверяет соответствующие атомы с точки зрения ограничений на число месяцев в году и дней в месяце.

Существуют шаблоны, охватывающие различные способы записи имен людей, включающие фамилию с инициалами, а именно:

Фамилия И. О. (с пробелом между инициалами и без)

Фамилия И.

И. О. Фамилия (с пробелом между инициалами и между инициалом и фамилией или без)

И. Фамилия (с пробелом между инициалом и фамилией или без).

Процедура атрибуции производится применительно к слову, написанному полностью (фамилии), а начальные буквы имени и отчества отсекаются.

Необходимо заметить, что при полном написании компонентов Ф. И. О. любым из следующих способов:

Фамилия Имя Отчество

Имя Отчество Фамилия

Фамилия Имя

Имя Фамилия

Имя Отчество, —

сборка соответствующего комплекса осуществляется только на этапе синтаксического анализа. Напомним, что с графематической точки зрения каждая составляющая является отдельным токеном и предварительно характеризуется как имя собственное. На этапе атрибуции эти гипотезы проверяются по словарю словоформ. Аналогичным образом обстоит дело с составными наименованиями географических объектов (и не только), ср. *Нижний Новгород*, *Черное море*, *Петр Первый* (подробнее о различных сложностях, возникающих при морфоанализе имен собственных, в частности имен людей, см. [Кобзарева 2004]).

Обработка аббревиатуры с окончанием, будь то с дефисом или без (ср. *КВНы*, *КВН-ах*), предполагает вначале отсечение строчных букв и поиск той части токена, что набрана заглавными буквами (*КВН*), в словаре аббревиатур. Если она найдена, мы получаем лемму опорного слова аббревиатуры (в данном случае — *клуб*) и ее графематический род. Далее имеющееся окончание сравнивается с окончаниями существительных данного рода на предмет совпадения. В случае успеха мы получаем (дополнительно к сведениям, пришедшим из словаря аббревиатур) морфологические характеристики токена по словоизменительным категориям (числа и падежа).

Схожей проверке подвергаются токены, состоящие из записанного цифрами числа и буквенного окончания типа *10-е*, *3-мя*, *126-ти* — для них производится поиск соответствующего окончания (точнее, псевдоокончания) по всем парадигмам количественных

и порядковых числительных. В итоге нередко получается несколько вариантов интерпретации. Например, токен *20-е* атрибутируется как порядковое числительное единственного числа, среднего рода, в именительном либо винительном падежа. А такие токены, как, скажем, *31-й* и *31-м*, неоднозначны даже с точки зрения того, количественные это числительные или порядковые, и индекс омонимии у них еще выше.

Если в тексте встречается цифровая запись числа без буквенного окончания, она может обозначать как количественное, так и порядковое числительное. До проведения синтаксического анализа невозможно квалифицировать сочетания типа *2 группы*, *137 км* как выражения, обозначающие количество предметов или их порядок при счете. На этапе атрибуции токенов обработка цифровой записи предполагает отделение с правого конца цифр, выражающих числа от 0 до 20, и приписывание двух лемм, соответственно для количественного и порядкового числительных¹¹. Например:

31	числ. колич.	один
31	числ. порядк.	первый
610	числ. колич.	десять
610	числ. порядк.	десятый
1113	числ. колич.	тринадцать
1113	числ. порядк.	тринадцатый

* * *

Возвращаясь к нашему опыту атрибуции буквенных токенов, следует заметить, что существуют разные методы проведения морфологического анализа. Предлагаемый подход, опирающийся на использование словаря словоформ, в настоящее время является наиболее распространенным, так как существенно экономит усилия лингвистов и разработчиков программного обеспечения. Такое «декларативное» обеспечение морфологического анализа, однако, имеет известный недостаток, а именно неспособность справляться со словами, которых нет в словаре (именами собственными, неологизмами, разнообразными дериватами, намеренно или случайно искаженными словами). Частично эта проблема может быть решена

¹¹ Это решение обусловлено дальнейшей задачей связывания токенов в ходе конструктивно-синтаксического анализа. Соответствие между членами количественных групп справедливо квалифицируется Л. Л. Иомдиным [1990: 62] как одно из самых сложных и запутанных в русском синтаксисе.

при помощи механизма морфологического предсказания, однако он не всегда оказывается эффективным.

В качестве альтернативы мы рассматривали возможность разработки процедурного морфологического анализа, направленного на выявление морфемного состава словоформ. Разработка такого анализатора потребовала бы колоссальной подготовительной работы, но зато дала бы возможность обрабатывать слова, не идентифицированные по словарю. Эта идея казалась заманчивой нам еще и потому, что она была бы созвучна процедуре анализа текста на других этапах. Поясним сказанное.

Мы уже видели, что при графематическом анализе алгоритм работает по принципу постепенного, многоступенчатого наращивания единиц слева направо: от символов к атомам, от атомов к токенам, с последующей сборкой частей предложения, предложения и абзаца. Процедура опирается на словарь элементарных сегментов (символов) и правила интерпретации, разработанные для каждого шага процедуры связывания соседних сегментов между собой. Та же идеология лежит в основе конструктивно-синтаксического анализа, о котором речь пойдет ниже. С этой точки зрения было бы логично, если бы морфологический анализ строился на основе словаря морфем с приписанными им значениями признаков, обуславливающими возможность или невозможность соединения морфем друг с другом. Это способствовало бы более успешной обработке новых слов и дериватов вне зависимости от степени их закреплённости в языке, позволило бы делать предсказания относительно их семантики. Кроме того, мы могли бы существенно сжать словарь словоформ, в частности за счет многочисленных глагольных приставочных образований. Но, пожалуй, основная привлекательность этого подхода лежит в плоскости методологии: выбор такого способа морфологического анализа позволил бы говорить об универсальном алгоритме обработки языка на разных этапах — от графематики до синтаксиса.

В то же время очевидно, что создание морфемного анализатора с нуля — это весьма трудоемкое мероприятие, по сложности не уступающее синтаксическому анализу. Потребность в системном описании морфем русского языка, учитывающем их формальные варианты (алломорфы), значения и комбинаторные свойства была осознана более трех десятилетий тому назад [Милославский 1980], но само описание до сих пор не разработано. Имеющиеся морфемные и словообразовательные словари (ср. (Тихонов 1985; Кузнецова, Ефремова 1986; Ефремова 2000)) не решают эту комплексную задачу

в полном объеме. Но даже если существовал бы толковый словарь морфем, учитывающий не только присущие им значения, но и возникающие в различных окружениях семантические «приращения», автоматический разбор многоморфемных слов также не был бы простой задачей. Здесь, как и на других этапах анализа, осложняющим фактором является много-многозначное соответствие между формой и значением.

Что же касается идеи об универсальном алгоритме обработки языка на разных этапах, она, по-видимому, все равно не смогла бы быть воплощена. Ведь в языке, помимо словоформ, есть токены другой природы, требующие совсем иных подходов к интерпретации. Если говорить не о традиционном морфоанализе, а в целом об атрибуции всех токенов в тексте, данная процедура в принципе не может быть универсальной и последовательной — это обусловлено спецификой текстов на ЕЯ.

Глава 3

КОНСТРУКТИВНО-СИНТАКСИЧЕСКИЙ АНАЛИЗ ПРЕДЛОЖЕНИЯ

СЕМАНТИКО-СИНТАКСИЧЕСКИЙ СЛОВАРЬ

Понятие семантико-синтаксического варианта

Конструктивно-синтаксический анализ¹ предложения начинается с обращения к семантико-синтаксическому словарю с целью извлечения соответствующей информации в отношении входящих в предложение буквенных токенов. В итоге к значениям графических и морфологических признаков, приписанных им на предшествующих этапах, добавляется соответствующая семантико-синтаксическая информация.

В описываемом словаре лексические единицы представлены в виде набора семантико-синтаксических вариантов (ССВ). ССВ — это, по сути, аналог значения (или оттенка значения) в толковом словаре, с той разницей, что существенное внимание при выделении этих вариантов уделяется их синтаксической реализации. Вследствие этого набор семантико-синтаксических вариантов слова, как правило, отличается от совокупности его значений и оттенков значений, выделяемых в традиционных словарях.

С одной стороны, то, что в толковом словаре выступает в качестве единого значения, в нашем описании может быть разбито на несколько ССВ. Чаще всего это обусловлено тем, что в традиционных толковых словарях характеристика подчинительных связей слова дается общим списком. К примеру, глагол *просить* в значении ‘до-

¹ Вслед за И. П. Распоповым [1970: 31], М. В. Никитиным [2007: 503] и Я. Г. Тестельцом [2001: 438] мы считаем целесообразным рассматривать синтаксис как многослойное образование и выделяем «конструктивный синтаксис» и «коммуникативный синтаксис». При этом, в отличие от М. В. Никитина [2007: 504], мы не ставим знак равенства между конструктивным и формальным синтаксисом: конструктивно-синтаксический анализ в описываемом лингвистическом процессе задействует информацию о значении лексических единиц из специально разработанного семантико-синтаксического словаря, которому и посвящена настоящая глава.

бываться чего-л. у кого-л.; склонять к исполнению желаемого' имеет следующие пометы: «(кого-что), кого-чего, о чем, с инф., с придат. предл.». Понятно, что такое перечисление едва ли пригодно для автоматической обработки языка. Требуется преобразовать этот список в структурированное описание валентностей глагола и способов их реализации. Глагол *просить* имеет три валентности: субъекта, контрагента и объекта. Субъект и контрагент имеют стандартное выражение (соответственно подлежащим и прямым дополнением), в то время как валентность объекта (содержание просьбы) может быть реализована существительным в родительном падеже (*хлеба, слова*), предложно-падежной формой «о S_{предл.}» (*о помощи*), инфинитивом (*помочь*) либо изъяснительным придаточным, вводимым союзом *чтобы* (*чтобы помогли*). В итоге одно словарное значение в нашем словаре представлено в виде четырех ССВ.

С другой стороны, количество ССВ, выделяемых у той или иной леммы, может быть меньше, чем число словарных значений. Во-первых, есть случаи, когда правомерность выделения того или иного значения в толковом словаре вызывает сомнения. Например, у прилагательного *светлый* Большой толковый словарь русского языка (БТС 2009), вслед за Словарем русского языка в четырех томах (СРЯ 1981–1984), выделяет значение 'блестящий, излучающий блеск', которое иллюстрируется, в частности, речением *светлый кинжал* (в четырехтомном словаре — цитатой из стихотворения М. Ю. Лермонтова «Кинжал», которая была преобразована в речение). С нашей точки зрения, отдельного значения здесь нет (ср. ?*светлое лезвие*, ?*светлый нож*, ?*светлые украшения*), есть лишь авторское окказиональное употребление.

Во-вторых, семантико-синтаксический словарь более ориентирован на описание денотата слова, чем его сигнификата. Такой выбор неизбежен в контексте автоматической обработки языка. Напротив, традиционная лексикография при описании семантики слова руководствуется общностью сигнификативного компонента отдельных значений. Различие между этими подходами отчетливо видно на примере многозначных прилагательных типа *тяжелый, мягкий, темный, высокий* и др. Вот как словарь² описывает семантику слова *высокий* (конец словарной статьи, содержащий фразеологические обороты и производные слова, опущен):

² Здесь и далее источником лексикографической информации служит (БТС 2009).

1. Большой по протяженности снизу вверх; находящийся далековерху; превосходящий обычную, среднюю высоту (противоп.: низкий). *В-ая гора. В-ие облака. В-ая трава. В. каблук. В. потолок. В. прыжок. В. лоб.*
2. Значительный по количеству, интенсивности, степени проявления и т. п.; превышающий средний уровень, среднюю норму. *В-ая активность. В-ая производительность труда. В. урожай. В-ие темпы. В-ие цены. В-ое напряжение, давление. В-ая концентрация вещества. В-ая чувствительность. В. уровень жизни. В-ие требования. В-ое мастерство.*
3. Очень хороший; отличный. *В-ие оценки. В-ие достоинства. Показать в. класс. Товары высокого качества. Быть высокого мнения о себе. Высокой пробы кто-, что-л.*
4. Выдающийся по значению; почетный, важный. *В-ая награда. В. пост. В-ая честь. В-ое звание. В-ое общественное положение. Оправдать в-ое доверие. Приветствовать высокого гостя. Вращаться в высоких сферах. В-ие договаривающиеся стороны.*
5. Возвышенный, благородный, лишенный корыстных помыслов. *В. порыв. В-ое чувство. В-ое искусство. В. идеал. В-ое предназначение. В-ая цель. Говорить о высоких материях (обычно ирон.).*
6. Приподнятый, торжественный (о языке, стиле). *В. слог.*
7. Вызываемый колебаниями большой частоты; тонкий и звонкий (о звуке, голосе). *В-ая нота. В. дискант.*

Из этих значений лишь шестое и седьмое денотативно ориентированы и без труда выделяются в семантико-синтаксическом словаре в качестве отдельных вариантов. Описание остальных значений с точки зрения семантических требований к управляющему существительному затруднено. Как видно из примеров, в первом значении (и соответствующем оттенке) прилагательное *высокий* может сочетаться и с конкретными, и с абстрактными существительными. Что касается второго, третьего и пятого значений, все предложения демонстрируют сочетаемость с абстрактными существительными, причем без возможности дальнейшей специализации, а если специализация невозможна, то нет и самостоятельных ССВ. Наконец, четвертое значение, как и первое, не предполагает никаких, даже самых грубых (в терминах «конкретное — абстрактное»), ограничений на тип существительного.

Подводя итог, мы констатируем, что число ССВ, выделяемых у того или иного слова, как правило, не совпадает с количеством его словарных значений.

***От морфологической атрибуции словоформы
к ее семантико-синтаксическому варианту***

Как известно, «в последние десятилетия проблема многозначности наиболее остро стоит в компьютерной лингвистике» [Иомдин 2014: 89]. Основное назначение семантико-синтаксического словаря состоит в том, чтобы способствовать снятию лексико-грамматической неоднозначности: графических и морфологических характеристик, приписанных токенам на предыдущих этапах анализа, для этого, как правило, недостаточно. Благодаря специально разработанному формату описания, словарь представляет в структурированном виде информацию об особенностях реализации слова (точнее, каждого из его ССВ). Сравнение словарной информации с текстовыми данными позволяет снизить (а зачастую и снять) грамматическую омонимию словоформы в тексте, а также сократить ее лексическую многозначность, уменьшив количество допустимых интерпретаций.

Как организована информация в данном словаре? Выше было сказано, что словарная статья состоит из списка ССВ. При каждом из этих вариантов имеется фильтр, содержащий графические и морфологические требования к его реализации (фильтр может включать несколько ограничений). Например: «с прописной буквы», «только 3 л.», «только полн.». Если никаких графических и морфологических ограничений для данного варианта нет, фильтр пуст. Задача фильтра заключается в том, чтобы в ходе синтаксического анализа отсеивать те ССВ, которые не соответствуют информации, приписанной соответствующему токеноу на предшествующих этапах анализа.

Более подробно эта процедура может быть представлена следующим образом. Как правило, словоформа имеет несколько вариантов морфологической атрибуции (а иногда и несколько лемм). В семантико-синтаксическом словаре по лемме ищется соответствующая словарная статья, представляющая собой список ССВ. Если лемма не одна, берутся все ССВ для всех лемм. Далее нужно «примерить» каждый «морфологический вариант» на каждый «семантический вариант». Происходит последовательное сопоставление морфологической и графической информации, приписанной каждому из результатов атрибуции, с фильтром каждого из ССВ. Как правило, в ряде случаев такое сравнение выявляет противоречия. Следовательно, итоговое число вариантов с приписанной им графической, морфологической и семантико-синтаксической информацией обычно оказывается меньше того, что может быть получено сугубо комбинатор-

но. В целом, однако, вариативность на этом этапе резко возрастает; ее снижение происходит позднее — при выявлении синтаксических связей между токенами и их объединении в словосочетания.

Поясним сказанное на примере. Допустим, в тексте встрети́лась словоформа *леса*. Результат атрибуции данного токена выглядит следующим образом:

лес	1	сущ; неод; мр	род; ед
лес	2	сущ; неод; мр	им; мн
лес	2	сущ; неод; мр	вин; мн
леса	1	сущ; неод; жр	им; ед
леса	2	сущ; неод; жр	им; ед ³
леса	2	сущ; мн ⁴ ; неод	им; мн
леса	2	сущ; мн; неод	вин; мн

Итак, у нас есть семь вариантов морфологической атрибуции. Обратимся теперь к словарю. Совокупно по леммам *лес* и *леса* мы находим пять ССВ, смысл которых для простоты изложения мы сформулируем при помощи обычных словарных дефиниций, ср.:

1. Множество дикорастущих деревьев, расположенных на большом пространстве; пространство, обильно поросшее деревьями.
2. Большое количество, множество (о возвышающихся предметах).
3. (собир.) Срубленные деревья как строительный и промышленный материал; бревна.
4. То же, что *леска*.
5. (мн) Временное сооружение, устанавливаемое у стен здания для строительных и ремонтных работ.

Как видно из помет (которые, напомним, в семантико-синтаксическом словаре входят в состав фильтра), третий вариант может реализоваться только при условии, что слово *лес* стоит в единственном числе, а пятый, — напротив, ограничен множественным числом, так как слово *леса* в данном значении принадлежит к существительным *pluralia tantum*.

³ В значении 'леска' данная словоформа имеет вариативное ударение.

⁴ Включение «мн» в перечень неизменяемых морфологических характеристик (с дублированием этой пометы в значениях словоизменительных категорий) — способ обозначения существительных *pluralia tantum*, «унаследованный» от проекта АОТ.

Осталось наложить семантико-синтаксические варианты на результаты морфоанализа, принимая во внимание фильтры. В итоге остаются девять вариантов: пять — для леммы *лес* (три единственного числа и два множественного числа), два — для единственного числа существительного женского рода *леса* (с вариантами ударения) и два — для существительного *pluralia tantum леса*. Как и ожидалось, в результате использования семантико-синтаксического словаря число вариантов возросло, однако увеличение оказалось не столь значительным (с семи до девяти), что объясняется наличием грамматических ограничений у двух из пяти значений.

Условия реализации

Продолжим рассмотрение того, как устроено описание лексики в семантико-синтаксическом словаре. Прежде всего следует подчеркнуть, что в нем нет привычных словарных дефиниций наподобие тех, что были приведены в примере выше для удобства изложения. Вместо этого используется специально разработанный формат описания, состоящий из двух частей — «условий реализации» и «собственного значения». В совокупности эти два компонента позволяют осуществлять такое семантическое описание, которое выявляет не только глубинные аспекты значения лексической единицы, но и совокупность семантических и грамматических требований к сочетающимся с ней словам.

Условия реализации слова включают указание на его частеречную принадлежность и комплексное описание семантических и грамматических требований к другим словам, способным сочетаться с ним в синтаксических конструкциях (прежде всего — в рамках сильного управления). Комплексное описание может включать вплоть до восьми условий, каждое из которых представляет собой группу признаков со своими значениями. Для разных частей речи условия реализации имеют различное наполнение. Наиболее сложными со структурной точки зрения являются условия реализации у глаголов — для них предусмотрена возможность заполнения всех восьми групп признаков (для потенциальных восьми актантов⁵).

⁵ Указанное число является произвольным и взято «с запасом». С одной стороны, наибольшее число валентностей у глаголов в русском языке, отмеченное у глагола *командировать*, равняется шести [Апресян 1995а: 137], с другой стороны, в ряде случаев граница между обязательными и факультативными валент-

Если у слова выделяется несколько ССВ, то каждый из них описывается через присущие ему условия реализации и собственное значение. Обоснованность выделения того или иного ССВ обеспечивается прежде всего спецификой его условий реализации.

Условия реализации, как и фильтр (см. выше), направлены на то, чтобы сократить лексико-грамматическую неоднозначность. Но, в отличие от фильтра, они ориентированы не «назад» (к уже имеющейся графической и морфологической информации), а «вперед» — к тому, что выявляется в ходе синтаксического анализа. К примеру, наречие в русском языке может быть синтаксически связано с глаголом, прилагательным, наречием или существительным, быть способным или неспособным выступать в функции предикатива, иметь или не иметь предпочтительную (или даже закрепленную) позицию по отношению к главному слову, причем эта позиция может располагаться как перед главным словом (*очень красивый, глубоко порядочный*), так и после него (*форель по-царски, дорога домой*). Сопоставление записанных в словаре условий реализации конкретного наречия с текстовыми данными позволяет отбросить неподходящие варианты и в итоге сократить число возможных интерпретаций.

Рассмотрим по очереди структуру условий реализации для слов, принадлежащих к разным частям речи и лексико-грамматическим разрядам. Как уже говорилось, для глагольных ССВ предусмотрена возможность заполнения восьми групп признаков — соответственно для восьми потенциальных актантов в тексте⁶. Каждый актант характеризуется следующими признаками: «тип синтаксического заполнения» (с такими, например, значениями, как ‘именная группа в родительном падеже без предлога’, ‘инфинитив’, ‘придаточное предложение, вводимое союзом *чтобы*’), «тип семантического заполнения» (с семантическими признаками: ‘человек’, ‘транспортное средство’, ‘абстрактная сущность’, ‘контейнер’ и пр.), а также «семантическая валентность» (субъект, инструмент, начальная

ностями неоднозначна (см., например, [Апресян 1995а: 119–163; Семиотика и информатика 1998; Тестелец 2001: 156–228]).

⁶ Если та или иная семантическая валентность глагола является нереализуемой (например, инкорпорированной), она не указывается. Это решение имеет свои недостатки, связанные, в частности, с возможностью нереализуемых валентностей «выходить на поверхность» на более обширных отрезках текста. В дальнейшем планируется усложнить формат описания, сделав его более «семантически ориентированным», что позволит последовательнее отражать диатезы глагола и различные актантные деривации (см. [Тестелец 2001: 411–436]).

точка / конечная точка и т. д.). Заметим, что в семантико-синтаксическом словаре понятие видовой пары не используется, и каждый глагол описывается отдельно в совокупности всех его ССВ. Это решение обусловлено разнородностью видовых пар (разным семантическим соотношением между соответствующими глаголами несовершенного и совершенного вида), а также тем, что их члены нередко обладают разными системами лексических значений⁷. Соответственно, причастия и деепричастия «привязываются» к конкретному глаголу.

Все компоненты этого описания значимы для дифференциации ССВ глагола: при автоматическом анализе предложения они способствуют его правильной интерпретации. Хорошо известно, что тождество синтаксической конструкции не означает тождества реализуемой валентности. К примеру, глаголы *гордиться* и *забивать* (в знач.: ‘ударами вбить, вогнать во что-л.’) управляют существительным в творительном падеже без предлога, при этом реализуя разные валентности, соответственно, их описание будет различаться, ср.:

гордиться (условия реализации) глагол; первый актант: <...>; второй актант: падеж = творительный / предлог = без предлога / сем. признаки: неспецифицир. / валентность = объект

забивать (условия реализации) глагол; первый актант: <...>; второй актант: <...>; третий актант: падеж = творительный / предлог = без предлога / сем. признаки: базовый = предмет; агрегат. сост. = твердый / валентность = инструмент

Более того, одна и та же предложно-падежная форма при наличии единственного глагола в предложении также не гарантирует тождества реализуемой валентности, ср. *пить чай с вареньем / с женой / с удовольствием*. Впрочем, в данном случае мы имеем дело с так называемым слабым управлением, которое не входит в актантную

⁷ Например, у глагола *встречаться* есть два значения (‘время от времени иметь свидания, быть в близких отношениях’ и ‘бывать, попадаться’), отсутствующие у глагола *встретиться*; у глагола *обвинять* есть значение ‘поддерживать обвинение, выступать в качестве судебного обвинителя, прокурора’, которого нет у соотносительного по виду глагола *обвинить*, и т. д. Помимо этого, видовые соотношения у многозначных глаголов нередко формируются на уровне их отдельных значений, ср. *мочить* — *намочить*, *замочить* (‘делать мокрым, влажным’) и *вымочить* (‘пропитывать влагой для придания каких-л. свойств’). Такое положение вещей диктует необходимость создания индивидуального описания для каждого глагола.

рамку глагола и не влияет на его значение: во всех трех предложениях реализуется один и тот же ССВ глагола *пить*.

Однако нередко бывает так, что разные ССВ глагола характеризуются одинаковым набором валентностей, каждая из которых выражается регулярным образом (к примеру, субъект — именной группой в именительном падеже, объект — прямым дополнением и т. д.). Иными словами, мы имеем и тождественные наборы семантических валентностей, и тождественные синтаксические реализации этих валентностей. Тогда единственный способ их разграничить — это учитывать семантические признаки и их значения⁸. К примеру, глагол *жарить* в трех (из четырех) значений имеет валентности субъекта и объекта, ср.:

- 1) что. Приготавливать (мясо, рыбу, картофель и т. п.), подвергая действию жара без воды. *Ж. котлеты. Ж. картошку на масле;*
- 2) кого-что. Разг. Обжигать лучами, палить (о солнце). *Солнце жарит нещадно. □ безл. Ну и жарит сегодня! Спину так и жарит!*
- 3) что. Разг. Жарко топить (печку, баню и т. п.). *Зачем целый день ж. печку? Уже и так жарко.*

Несложно заметить, что семантические валентности у глаголов в данных ССВ заполняются существительными с разными значениями семантических признаков. Убрав (для краткости) из условий реализации сведения о валентностях и их синтаксической реализации, получим:

- 1) (условия реализации) глагол; первый актант: сем. признаки: базовый: живое существо; человек = +⁹; второй актант: сем. признаки: базовый: предмет; продукты питания = +;
- 2) (условия реализации) глагол; первый актант: сем. признаки: базовый: предмет; второй актант: сем. признаки: базовый: живое существо / часть тела;

⁸ Такая ситуация характерна, в частности, для глаголов физического действия. Так, в [Скребцова 1996] было показано, что изменение семантического типа субъекта и/или объекта (по отношению к ситуации, описываемой основным значением соответствующего глагола) является важным индикатором его употребления в производном значении.

⁹ Здесь и далее некоторые семантические признаки обозначаются фактически через их значения, а в качестве значения стоит «+».

- 3) (условия реализации) глагол; первый актант: сем. признаки: базовый: живое существо; человек = +; второй актант: сем. признаки: базовый: предмет.

Прилагательные также снабжаются семантической характеристикой существительных, способных с ними сочетаться, что позволяет в ходе синтаксического анализа сокращать лексическую неоднозначность. Так, прилагательное *жесткий* в своем прямом значении согласуется с предметными существительными (*кровать, корка* и т.д.), а в производных — с абстрактными (*характер, тон; срок, требования*). Дальнейшее разграничение производных значений ('суровый, резкий' vs. 'не допускающий отклонений') требует более тонкой «настройки» ожидаемых значений признаков.

Актантной структурой обладают также отглагольные имена со значением действия (*переезд, раздача, отправление*) и некоторые существительные, образованные от прилагательных (*возможность, доступность, синева*). Помимо таких очевидных номинализаций, есть целый ряд других существительных, которые, как правило, имеют при себе зависимые слова, ср. *цена, доход, разница, довод, отец, вера, жажда, ревность*. При выделении подобных существительных мы опираемся не столько на отношения производности, сколько на данные толковых словарей (сведения о подчинительных связях слова, дефиницию, речения и цитаты), а также на материалы НКРЯ.

У многих предметных существительных (точнее, их ССВ) условия реализации являются пустыми, т.е. не содержат ни одной характеристики. Эта ситуация характерна для основных значений таких существительных, как *стол, дом, синица, сосна, чашка, автомобиль* и пр., где отчетливо преобладает денотативный компонент значения, а производные реализуются в условиях синтаксических ограничений. Так, у существительных *стол* и *дом* есть производные значения, связанные с названием различного рода учреждений, ср. *стол заказов (находок), Дом творчества (культуры, архитектора, офицеров* и др.). В толковых словарях эти значения сопровождаются грамматической пометой «чего» или «кого/чего». В нашем семантико-синтаксическом словаре эта информация отражена в условии реализации, где указано, что данный ССВ требует зависимую именную группу в форме родительного падежа без предлога с определенными семантическими признаками и их значениями. (Справедливости ради следует сказать, что слово *дом* в указанном значении (и только в нем) пишется с заглавной буквы, поэтому выбор нуж-

ного ССВ может произойти за счет информации, содержащейся в фильтре, — правда, при условии, что данное слово не занимает инициальную позицию в предложении или прямой речи.)

Для предлогов в условиях реализации заполняются падеж управляемого существительного и значения его семантических признаков. Благодаря такому описанию наглядно разграничиваются значения предлога *в*, реализующиеся в следующих предложно-падежных формах:

в <июне>: (условия реализации) предлог; падеж = предложный / сем. признаки: базовый: время

в <лесу>: (условия реализации) предлог; падеж = предложный / сем. признаки: базовый: предмет; пространств. = +

в <лес>: (условия реализации) предлог; падеж = винительный / сем. признаки: базовый: предмет; пространств. = +

Составные предлоги (которые ранее — в ходе графематического анализа — были найдены в словаре устойчивых словосочетаний и соответственно помечены) также могут иметь более одного значения, ср.:

в течение <июня>: (условия реализации) предлог; падеж = родительный / сем. признаки: базовый: время

в течение <лекции>: (условия реализации) предлог; падеж = родительный / сем. признаки: базовый: действие

Для словосочетаний, которые могут функционировать и в качестве свободных, и в качестве связанных (ср. *до лампочки*, *по барабану*), предусмотрен учет обоих вариантов в ходе синтаксического анализа. В первом случае из семантико-синтаксического словаря будет получена информация, касающаяся отдельно предлога и существительного, во втором — информация, связанная с устойчивым сочетанием.

Для ССВ наречий указывается частеречная характеристика главного слова и значения его семантических признаков, а также фиксированность / нефиксированность позиции наречия по отношению к главному слову (и если позиция фиксирована, то какая — препозиция или постпозиция), способность / неспособность выступать в предикативной функции.

В отношении других частей речи (местоимений, числительных, союзов, частиц, вводных слов и оборотов) следует сказать, что, помимо снятия лексико-грамматической омонимии внутри предложе-

ния, они играют важную роль в обеспечении локальной и глобальной когеренции текста. Соответственно, наряду с описанием в семантико-синтаксическом словаре, они также рассматриваются на этапе коммуникативного синтаксиса, подразумевающим учет структурных, прагматических, риторических особенностей построения текста; об этом речь пойдет в следующей главе.

Собственное значение

Обратимся теперь ко второму компоненту описания семантико-синтаксических вариантов — к их «собственному значению». Оно складывается из значений соответствующих семантических признаков.

Заметим, что выше речь шла об использовании семантических признаков в условиях реализации, где они служили для характеристики слов, способных выступать в сочетании с тем или иным ССВ данного слова. Тот же самый набор семантических признаков с их значениями задействуется также и для описания самих ССВ как таковых, отдельно от их контекстуальных возможностей.

При разработке списка семантических признаков и их значений мы учитывали результаты, опубликованные в ряде источников, в том числе Русский семантический словарь (1998–2007), словари под редакцией Л. Г. Бабенко (БТСРГ 2007; БТСРС 2008), электронную онтологию проекта UNL¹⁰, список лексико-семантических классов в Национальном корпусе русского языка¹¹, отдельные идеографические словари (Hallig, Wartburg 1963; Баранов 1995) и некоторые другие. Прежде всего мы выделяем так называемый базовый семантический признак (со значениями ‘живое существо’, ‘предмет’, ‘время’, ‘действие’, ‘качество’, ‘количество’ и др.), благодаря которому осуществляется самая приблизительная, грубая характеристика ССВ.

Дополнительные семантические признаки формально не подчинены базовым, поскольку в ряде случаев это удобно. К примеру, лошадь, будучи животным, одновременно может служить и транспортным средством, однако в иерархически выстроенной онтологии транспортное средство рассматривалось бы как тип предмета. Таким образом, пришлось бы выбирать между указанием на лошадь как живое существо и на лошадь как транспортное средство. Наше описание преодолевает необходимость подобного выбора, ср.:

¹⁰ См.: <http://www.undl.org/unlsys/uw/UNLOntology.html>

¹¹ См.: <http://www.ruscorpora.ru/corpora-sem.html>

лошадь: (словарное значение: 'крупное домашнее животное, используемое для перевозки людей, грузов и т. п.') сем. признаки: базовый: животное = +; транспортное средство = +.

Однако, по сути, дополнительные семантические признаки, как правило, производны от базовых. Таковы, к примеру, признаки «агрегатное состояние» для предмета, «цвет» для качества, «физическое действие» для действия и т. д. В то же время есть и независимые признаки, ср. «наличие наблюдателя» или «релятивность», о которых речь ниже.

В некоторых случаях в качестве дополнительных семантических признаков фигурируют высоковероятные, хотя и необязательные компоненты значения. Таковым, к примеру, является признак «ювелирное украшение» в семантике слова *безделушка*, ср.:

безделушка: (собственное значение) сем. признаки: базовый: предмет; ювелирное украшение = +.

Следует подчеркнуть, что выделенные семантические признаки не имеют однозначной соотнесенности с частями речи или лексико-грамматическими разрядами. Так, значением 'действие' по базовому семантическому признаку помечаются ССВ не только глаголов, причастий и деепричастий, но и отглагольных существительных, а также образованных от причастий прилагательных (*усталый, жареный, растворимый*) и существительных (*раненый, заведующий, слабовидящий*). Это необходимо в дальнейшем для анализа свернутых пропозиций.

Приписывание значения 'действие' осуществлялось преимущественно исходя из его словарной дефиниции, точнее, из ее опорного слова. Если опорное слово дефиниции характеризуется выраженной предикатной семантикой, то рассматриваемое слово помечается семантическим признаком 'действие'. К примеру, различные значения существительного *ревность* толкуются через слова *сомнение, боязнь, усердие*, прилагательного *усталый* — через причастия *ослабевший, истомленный, потерявший, истощенный* и т. д.

Важно отметить, что способность иметь при себе сильноуправляемые члены как таковая автоматически не влечет приписывания признака 'действие'. Например, значение слова *разница* 'величина, число, являющиеся разностью между двумя числами, величинами' не имеет отношения к действию, хотя регулярно реализуется вместе с зависимыми словоформами, ср.: *Р. между старым и новым окла-*

дом. Р. между прежней и настоящей ценой товаров. Р. в десять лет. Р. в сорок рублей. Десять лет разницы между кем-л. Сорок рублей разницы между чем-л. Р. во времени между Москвой и Красноярском.

В то же время следует признать, что опора на словарные определения не всегда дает однозначный и надежный результат. В ряде случаев грань между предметным и признаковым значением очень тонкая и не всегда адекватно фиксируется лексикографами. Применительно к словам, обозначающим атмосферные осадки, этот феномен был отмечен группой исследователей [Дюндик, Ковалева, Овсянников 1985], которые обратили внимание на то, что слово *снег* в одних контекстах имеет предметное значение ('белые хлопья, кристаллики льда'), а в других — событийное ('снегопад')¹². И хотя словарь разделяет эти два смысла, словарное толкование «событийного» значения слова *снег* отсылает к понятию «твердых атмосферных осадков», а слово *осадки* определяется как 'атмосферная влага, выпадающая на землю в виде дождя, снега, росы и т. п.' Круг замкнулся, и опорные слова определений *осадки*, *влага* ситуацию не прояснили.

Подобное положение вещей не ограничивается указанной семантической группой. Так, толкование слова *взятка* в первом значении ('деньги или материальные ценности, даваемые должностному лицу как подкуп, оплата действий в интересах дающего') охватывает и предметный, и предикатный смыслы (ср. *дать / получить взятку*, но *осужден за взятку*), которые, на наш взгляд, должны быть разведены (соответственно, в семантико-синтаксическом словаре у данного слова выделяются два ССВ). Таким образом, приписывание семантического признака 'действие' само по себе оказывается непростой исследовательской задачей.

Значение 'качество' (по базовому семантическому признаку) приписывается качественным прилагательным и образованным от них наречиям (точнее, их основным ССВ), а также относительным и притяжательным прилагательным в их производных ССВ, связанных с переходом в разряд качественных прилагательных (ср. *золотой характер, железная воля, каменное лицо, стальные нервы, театральный жест; медвежья походка, собачья жизнь*). Кроме того, им снабжаются существительные со значением признака, образованные от качественных прилагательных (*белизна, грубость*), а также

¹² Кроме того, слово *снег*, наряду с некоторыми другими существительными, даже в своем предметном значении может употребляться в функции событийного имени, ср. *Он не пришел из-за мозолей / грязи / снега* [Арутюнова 1976: 144–145].

оценочные существительные, способные выступать в предикативной функции, характеризуя лицо или вещь, ср. *умница, проходимец, растяпа, лопух, орел, дрянь, прелесть* и т. д.¹³

Значение 'предмет' (по базовому семантическому признаку) получают конкретные существительные — наименования неодушевленных предметов (в том числе собирательные) и веществ, а также относительные прилагательные (*золотые украшения, крестьянская изба, машинное масло*) и притяжательные прилагательные, перешедшие в разряд относительных (*лисий воротник, медвежья шуба*).

В теоретическом плане предложенный подход коррелирует с понятием функциональной транспозиции¹⁴ у Ш. Балли [1955: 130], а также с противопоставлением двух типов семантических отношений между словообразовательно связанными единицами — синтаксической и лексической деривации [Курилович 1962]. В нашем семантическом описании, как видно из приведенных выше примеров, слова, связанные отношениями синтаксической деривации, входят в один и тот же класс, поскольку это преобразование затрагивает лишь синтаксическую функцию производящего слова (вследствие изменения его частеречной принадлежности), но не его семантику (ср. *преподавать* → *преподавание, хитрый* → *хитрость, белый* → *белизна* и т. п.).

Опора на содержательные признаки является необходимым условием снятия (или сокращения) неоднозначности. Если однокоренные слова разных частей речи встречаются в тексте рядом, то, скорее всего, из набора присущих им ССВ реализуются именно те, что характеризуются одинаковыми значениями семантических признаков (случаи каламбуров довольно редки). Ср.:

Было установлено **золотое** содержание червонца — 7,74234 грамма чистого **золота**.

7 января я **переехал** в город; жена, ввиду стоявших сильных морозов, не решилась на **переезд** и осталась еще на неделю в Царском.

И красавица, и **умница, умней** моего дурака раз в десять.

Приведенные примеры, заимствованные из Национального корпуса русского языка, ограничены рамками одного предложения¹⁵,

¹³ Ср. [Грамматика Пор-Рояля 1991].

¹⁴ О направлениях транспозиции на материале русского языка см. [Богданов 1977: 157–169].

¹⁵ В силу специфики корпусных данных это замечание справедливо и в отношении последующих примеров.

однако нетрудно догадаться, что на более широком контексте схожее явление наблюдается еще чаще. Следовательно, совпадение значений семантических признаков можно использовать также на следующем этапе анализа для выявления когерентных связей в тексте.

* * *

Выше мы говорили о базовом семантическом признаке и его значениях. Обратимся теперь к тому, что можно совокупно назвать дополнительными семантическими признаками. С их помощью мы пытаемся зафиксировать разнообразные аспекты содержания того или иного ССВ.

Введение различных, не упорядоченных иерархически семантических признаков обеспечивает многомерность описания, при котором каждое слово может рассматриваться как совокупность семантических формул (ср. [Тузов 2003]). Помимо самостоятельной ценности, такой подход лучше справляется с задачей снятия омонимии. Это особенно актуально для существительных и глаголов, которые характеризуются в целом более высоким индексом неоднозначности, чем слова других частей речи. Целесообразность добавления тех или иных семантических признаков напрямую зависит от их способности снимать омонимию у возможно большего числа лексических единиц. Как правило, эта мера имеет также положительный эффект и для решения дальнейших задач в рамках анализа целостного текста.

Так, у существительных собственное значение включает ряд «нетривиальных семантических признаков» [Апресян 1980], имеющих двойное назначение: дифференциацию ССВ лексемы и выявление смысловых связей в тексте. Для решения первой задачи важен, например, такой признак, как «контейнер», позволяющий учитывать регулярный метонимический перенос у предметных существительных, обозначающих разного рода вместилища (ср. *стакан стоит на столе* vs. *выпил целый стакан*). Ввиду того что практически любой предмет может осмысляться в качестве вместилища, положительное значение данного признака приписывается только словам (точнее, их ССВ), обозначающим «классические» контейнеры типа *чашка, кастрюля, бочка, бутылка, чемодан, корзина* и т. п.¹⁶

¹⁶ Метонимия в целом представляет серьезную проблему для автоматической обработки языка. Даже регулярные метонимические переносы наподобие вышеупомянутого в словарях трактуются непоследовательно (игнорируются, упо-

Признак «множественность» актуален для существительных, у которых форма единственного числа обозначает множество людей или однородных предметов¹⁷. Частично дублируя морфологический признак собирательности (ср. *студенчество*, *детвора*, *воронье*, *листва*, *горох*), он не ограничивается существительными *singularia tantum* и позволяет охватить также такие имена, как *семья*, *полиция*, *команда*, *армия*, *полк*, *начальство*, *электорат*, *персонал*, *консилиум*, *флот* (в знач. 'совокупность судов'), *скот*, *свора*, *табун*, *лес*, *роща*, *школа* (в знач. 'учителя и/или ученики'), *поселок* (в знач. 'проживающие в нем люди'), *пачка*, *стонка*, *кипа*, *парк (машин)*, *библиотека* (в знач. 'собрание книг') и некоторые другие¹⁸. Многие объекты данного класса отличаются тем, что не имеют компактного расположения в пространстве или могут его утратить (например, *полк* солдат может быть расквартирован в деревне, *коллекция* картин — разделена между музеями, *табун* лошадей может разбежаться по степи, *стонка* бумаг — разлететься по полу и т. д.).

Еще один важный семантический признак для существительных связан со способностью объектов самостоятельно перемещаться и/или быть перемещаемыми. В ряде случаев это позволяет разграничить ССВ слова, ср. *лес* как 'массив деревьев' («перемещаемость» = -) vs. 'строительный материал' («перемещаемость» = +),

минаются в общей дефиниции через запятую, выделяются в качестве оттенков значения). Но охватить все «метонимические формулы» [Шмелев 1973: 224] вряд ли вообще возможно, учитывая их многообразие, а также различную степень закреплённости и контекстной обусловленности. Дополнительным осложняющим фактором являются окказиональные выражения, построенные по той или иной модели. Так, временное значение, реализуемое в сочетаниях *до чая*, *после армии* ('до чаепития', 'после службы в армии'), допускает различные прочтения в выражении *до телевизора* ('до покупки телевизора' или 'до просмотра телевизионных передач'). Ср. также выражение *до дяди* в приводимой Д. Н. Шмелевым [Шмелев 1973: 225–226] цитате из «Двух капитанов» В. Каверина, ср.: *Мы считали, что этих денег хватит до дяди, а там начнется увлекательная, превосходная жизнь*. В общем случае проблема заключается в том, что там, где конкретная лексика попадает в контексты, требующие имен с абстрактной семантикой, она должна получать событийное прочтение, однако содержащейся в предложении информации недостаточно для того, чтобы восстановить соответствующую пропозицию [Арутюнова 1976: 122–129; 142].

¹⁷ Ср. лексическую функцию Mult в модели «Смысл ↔ Текст» [Мельчук 1974: 88].

¹⁸ О структурных типах множества и их реализации в языке см., например, [Холодович 1979; Румянцева 2009].

земля как 'территория' («перемещаемость» = -) vs. 'почва, грунт' («перемещаемость» = +).

Для дальнейшего анализа текста представляется обоснованным ввести семантические признаки «выделенная часть» и «выделенное направление движения» (ср. [Talmy 1983]), характеризующие такие объекты, как шкаф, кресло, зеркало, дворец, автомобиль (и другие транспортные средства), очередь (см. также [Fillmore 1969; Всеволодова, Владимирский 2009: 208 и след.]). Они позволяют анализировать положения предметов в пространстве, описываемые выражениями типа *вертеться перед зеркалом, стоять за шкафом, левый флигель дворца* и т.д., что связано с более общей задачей интерпретации пространственного дейксиса. Показательно, что некоторые предложения, содержащие отсылки к объектам с выделенной частью или направлением движения, допускают двойное прочтение — дейктическое и недейктическое. В качестве иллюстрации сошлемся на пример Ю. Д. Апресяна [1995б: 635] *Девушка стояла перед машиной*, где при дейктической интерпретации девушка помещается между говорящим и машиной, а недейктическое прочтение учитывает наличие у автомобиля выделенного направления движения, и, соответственно, девушка стоит по ходу движения автомобиля, т.е. перед его капотом. При автоматическом анализе неоднозначность, присущая данному предложению, должна быть сохранена.

Данные признаки можно считать частными случаями признака «абсолютная/относительная пространственная ориентация» [Апресян 1980], ср. *верх (низ) шкафа*, но *верх (низ) кубика*. Абсолютную ориентацию по верху — низу имеют все объекты, обладающие выделенной частью, но не только они, ср. *дерево, гора*. Соответственно, объекты с выделенным направлением движения имеют абсолютную ориентацию по передней части — задней части; таковы, к примеру, транспортные средства.

Признак релятивности (ср. [Шмелев 1973: 135; Palmer 1982: 88]) призван отразить определенную дейктичность семантики слова, ее зависимость от точки отсчета. Положительное значение данного признака характерно для терминов родства, но не только для них (ср. *сосед, одноклассник, попутчик, предшественник*). В то же время для большинства наименований лиц (например, по профессии, ср. *механик, инженер, художник*) этот признак нерелевантен. Важно, что указанные два случая предполагают разные построения в информационной структуре (подробнее см. ниже).

Отметим также, что положительное значение данного признака сближает рассматриваемые существительные со словами других лексико-грамматических разрядов: личными и указательными местоимениями, наречиями, предложениями, частицами *вот* и *вон*, прилагательными типа *верхний* — *нижний*, *старший* — *младший* и другими лексическими единицами, семантика которых предполагает указание на ориентир (ср. понятие *landmark* в работе [Langacker 1988]). Очевидно, что, если бы мы в семантическом описании ограничивались только иерархически упорядоченной онтологической схемой, соответствующий компонент значения, равно как и указанная параллель, не были бы выявлены.

* * *

Большое внимание мы уделяем выделению семантических признаков для глагола, поскольку он является структурным и смысловым центром предложения¹⁹. Прежде всего, следует отметить признаки, связанные со способами глагольного действия. Руководствуясь практическими соображениями, мы использовали только те из них, которые способны служить для дифференциации значений у достаточно большого числа глаголов. Выделить такие семантические признаки и приписать их соответствующим ССВ глаголов не составляет большого труда, если способ действия представляет собой словообразовательный разряд (к тому же словарные дефиниции, как правило, строятся однотипно). В случае непоследовательно характеризованных способов действия эта задача требует ручной работы.

Перечислим некоторые способы глагольного действия, указание на которые включается в собственное значение глаголов:

- начинательный — *заговорить* ('начать говорить' vs. 'утомить разговором'), *заходить* ('начать ходить' vs. 'приходить куда-л.'), *пойти* ('начать идти' vs. 'вступить в брак');
- ограничительный — *поцарапать* ('царапать некоторое время' vs. 'оцарапать'), *пошить* ('шить некоторое время' vs. 'сшить');
- длительно-ограничительный — *проговорить* ('говорить некоторое время' vs. 'сказать'), *прогулять* ('гулять некоторое время' vs. 'не выйти на работу');
- накопительный — *налетать* ('пролететь в общей сложности какое-л. число часов, километров' vs. 'натолкнуться на прегра-

¹⁹ Более широко о регулярных семантических признаках глагола см. [Скребцова 2015a].

ду'), *сносить* ('износить' и vs. 'отнести куда-л. и принести обратно'), *вырезать* ('истребить резней' vs. 'удалить, извлечь'), *исходить* ('ходя, побывать во многих местах' vs. 'брать что-л. за основу');

- распределительный — *порубить* ('вырубить все, многое' vs. 'рубить некоторое время'²⁰), *перебить* ('убить всех, многих' vs. 'не дать договорить, прервать'), *перекатать* ('покатать всех, многих' vs. 'быстро переписать'), *переболеть* ('перенести много болезней' vs. 'перенести какую-л. болезнь');
- репродуктивный — *пересказать* ('рассказать заново' vs. 'рассказать все, многое'²¹), *пересушить* ('просушить заново' vs. 'сделать слишком сухим'²²);
- чрезмерно-нормативный — *перепечь* ('подвергнув более длительному или более сильному действию жара, лишить нужных вкусовых качеств' vs. 'испечь все, многое'²³), *перестоять* ('простоять дольше, чем нужно' vs. 'простоять где-л., пережидая').

Способы действия пронизывают систему глагольных значений в русском языке, вступая между собой в различные комбинации в семантике конкретных глаголов. Эффективность соответствующих признаков для семантико-синтаксического анализа очевидна.

Кроме указания на способ действия, для глаголов вводится признак «фаза» (со значениями 'начало'²⁴, 'продолжение', 'остановка', 'конец' и др.²⁵), который также способствует снятию неоднозначности у целого ряда глаголов. Так, семантическая структура глаголов типа *белеть*, *светлеть*, *сохнуть* характеризуется регулярной многозначностью: одно (или более) значений связано с идеей становления соответствующего свойства, а другое (другие) выражают наличие постоянного свойства, воспринимаемого неким наблюдателем. Следовательно, признак «фаза» (со значением 'становление') релевантен для первого типа значений, но не для второго.

²⁰ Ограничительный способ действия, см. выше.

²¹ Распределительный способ действия, см. выше.

²² Чрезмерно-нормативный способ действия, см. ниже.

²³ Распределительный способ действия, см. выше.

²⁴ Множество соответствующих глагольных значений в качестве своего подмножества содержит те ССВ, которые помечены признаком начинательного способа действия (см. выше).

²⁵ Понятие фазовости трактуется нами шире, чем это обычно принято, ср. также [Богданов 1985; Храковский 1987].

Характеризация глагольного значения по признаку «фаза» может оказаться полезной в дальнейшем при выявлении смысловых связей не только внутри предложения, но и текста. Возьмем в качестве примера следующую фразу:

Остановившийся было процесс разложения **возобновился**, **грозя** **со** всем развалить фронт, а с ним и Россию.

На фазу в этом примере указывают три токена: *остановившийся* и *возобновился* характеризуют стадии процесса *разложения*, в то время как деепричастие *грозя* реализует фазовое значение 'предбытие' применительно к событию, обозначенному глаголом *развалить* (*фронт*). Кроме того, данное деепричастие выражает причинно-следственную связь между пропозициями «процесс разложения возобновился» и «фронт разваливается».

В отношении деепричастия *грозя* следует также заметить, что оно (как и упоминавшиеся ранее глаголы *белеть*, *светлеть* и *сохнуть*) содержит в себе сему 'наблюдатель'²⁶. Вообще, эта сема присутствует у достаточно широкого пласта лексики (причем не только глагольной), связанной с восприятием качеств и признаков объектов (ср. *мелькать*, *шелестеть*, *сияние*, *горький* и т. д.). Выделение соответствующего семантического признака для характеристики глагольных ССВ обусловлено не столько его ролью в снятии лексической и синтаксической неоднозначности, сколько способностью восполнять имплицитные связи между пропозициями в тексте. Он позволяет связать информацию о воспринимаемом объекте с самим экспериенцером, который не входит в число актантов соответствующего глагола, но может употребляться в тексте при другом глаголе, ср.:

Там **мелькали** обольстительные призраки, но **он** не мог разглядеть их; **слышались** смешанные звуки — то голос славы, то любви: все это приводило **его** в сладкий трепет.

Еще один важный аспект глагольной семантики — признак «контролируемости / неконтролируемости» действия, позволяющий дифференцировать, к примеру, следующие ССВ глаголов²⁷, ср.:

²⁶ «Наблюдатель» понимается нами широко — как субъект не только перцепции, но и сознания (ср. [Падучева 2006]).

²⁷ Подробнее о контролируемости действий и отражении этого аспекта в языке см. [Зализняк 1992; Кустова 1992].

бренчать: ‘издавать звуки, подобные ровному глухому звону’ («контролируемость» = –) vs. ‘производить, вызывать такие звуки’ («контролируемость» = +, ср. *бренчать ключами*);

моргать: ‘непроизвольно быстро опускать и поднимать веки и ресницы’ («контролируемость» = –) vs. ‘подавать знак движением век’ («контролируемость» = +);

пройти: ‘оказаться принятым, зачисленным, избранным’ («контролируемость» = –, ср. *пройти по конкурсу*) vs. ‘идя, совершить путь куда-л.’ («контролируемость» = +).

Следует заметить, однако, что, будучи «скрытой» категорией [Филлмор 1981], контролируемость действия не обязательно выражается в том или ином контексте. Ср., например, следующие примеры из работы [Апресян 1995а: 176], где содержательная оппозиция контролируемости / неконтролируемости нейтрализуется: *Джон прострелил себе руку, Гусар загремел шпорами, Он разбил окно локтем, Больной моргнул* и т. д.

С точки зрения анализа целостного текста представляют интерес и другие семантические признаки, связанные с характером протекания действия, обозначенного глаголом. Так, по признаку, который можно назвать «предельность» или «целостность» (ср. [Виноградов 1947; Маслов 1959; Мелиг 1985; Телин 1985; Бондарко 1987]), всем глаголам совершенного вида может быть автоматически приписано значение ‘+’ (что, правда, в некотором роде дублирует уже имеющуюся информацию: если глагол совершенного вида, то он непременно предельный). Что же касается глаголов несовершенного вида, здесь требуется ручная работа по разделению тех из них, которые, по выражению Г. Г. Сильницкого [1983], характеризуются «потенциальной предельностью», и непредельных глаголов.

Другой кандидат в семантические признаки — понятие кратности, которое, как правило, рассматривается совокупно с понятиями длительности и интенсивности в рамках количественной аспектuality²⁸. В то же время иногда утверждается, что три указанных понятия представляют собой независимые семантические признаки, каждый из которых может реализоваться в виде отдельной категории [Храковский 1989: 11]. Это позволяет рассматривать кратность

²⁸ Ср.: «...в славянских языках количественная аспектuality воплощается в различных частных видовых значениях форм СВ и НСВ, выступают в качестве способов действия (многократные глаголы типа *хаживать, есть*), семельфактивные — типа *стукнуть, ахнуть* и т. д.)...» [Маслов 1978: 19].

вне контекста, связанного с глагольными способами действия (ср. [Бирюлин 1989]). Если взять за основу такой подход, то данный семантический признак может быть распространен на глаголы движения, традиционно описываемые в терминах оппозиции моторно-некратные vs. моторно-кратные (*бежать* — *бегать*, *вести* — *водить*), их многочисленные приставочные дериваты, а также глаголы, обозначающие колебательное движение (*болтаться*, *шататься*) (ср. [Апресян 1995а: 72]).

Что же касается других, более тонких аспектов глагольной семантики, заинтересованно обсуждавшихся в литературе после появления знаменитой классификации З. Вендлера (ср. признаки 'актуальность', 'наличие цели', 'мгновенность'), они, как справедливо замечает Х. Мелиг [1985], отличаются высокой зависимостью от контекста. Сложные переплетения функционально-семантических полей аспектуальности, темпоральности, залоговости, модальности, субъектности / объектности и др. в русском языке (см. [Бондарко, Буланин 1967; Пупынин 1991, 1996; Бондарко 1992; Храковский 1996]) не позволяют использовать эти признаки для характеристики глаголов как единиц языка.

* * *

У прилагательных выделяются параметрические ССВ, характеризующие предмет с точки зрения размера (величины, высоты, глубины, длины, ширины, толщины), температуры, веса. По значению «параметр = +» / «параметр = -» легко дифференцировать прямые и метафорические значения соответствующих прилагательных, ср.: *высокий дом* vs. *высокие отношения*, *широкая кровать* vs. *широкие взгляды*, *глубокое озеро* vs. *глубокий анализ* и т. п. Наличие в тексте прилагательного с положительным значением данного признака нередко служит сигналом скорого появления словосочетания с количественным числительным и, следовательно, оказывается полезным для выявления смысловых связей в предложении и тексте.

Применительно к прилагательным также заслуживает внимания предложенный Ю. Д. Апресяном [1995а: 66–67] семантический признак 'предельность'. Прилагательное является предельным, если оно может сочетаться с наречиями *абсолютно*, *совсем*, *совершенно*, *идеально*: к ним относятся прилагательные цвета, формы, параметрические прилагательные (*короткий*, *узкий*, *тонкий*, *холодный* и т. д.), а также прилагательные, обозначающие соответствие / несоответствие норме (типа *здоровый*, *хороший*). Присутствие в значении со-

ответствующей семы разграничивает, например, значения прилагательного *короткий* (ср. *совсем короткий шнур* vs. **совсем короткое замыкание*) и *красный* (ср. *абсолютно красный галстук* vs. **абсолютно красный воин*).

АЛГОРИТМ КОНСТРУКТИВНО-СИНТАКСИЧЕСКОГО АНАЛИЗА

Цель конструктивно-синтаксического анализа — выявление синтаксической структуры предложения. Как и на предыдущих этапах обработки текста, мы стремимся к тщательному учету линейного контекста, чтобы максимально сократить возможную неоднозначность. В соответствии с общим принципом работы лингвистического процессора это происходит исключительно на основе языковых данных, без привлечения статистических методов. Если имеющаяся информация не позволяет сделать надежный выбор в пользу какого-нибудь одного из альтернативных вариантов синтаксического разбора предложения, все они сохраняются — вплоть до следующего этапа анализа²⁹. Если какой-то элемент не удастся синтаксически связать ни с одним членом предложения, он временно (до следующего этапа) остается в «подвешенном» состоянии.

В методологическом плане нам близок подход В.А.Тузова, который формулирует свою исходную посылку следующим образом: «Поскольку компьютер не способен понимать тексты, а может их только выполнять, то остается практически один способ формализации языка — рассматривать предложение как запись последовательности выполнимых действий. Нравится нам это или нет, компьютер не оставляет нам другого выбора» [Тузов 2003: 4]. Исходя из этого единственно верным способом представления структуры предложения автору видится суперпозиция функций, ср.: «Предложение является алгоритмом, и никакой статический тип данных не адекватен ему. Поэтому, если какая-либо грамматика приписывает какому-то предложению нечто, отличное от суперпозиции (древовидную структуру, графовидную или нечто подобное), то такая грамматика не может быть адекватной естественному языку» [Тузов 2003: 429].

²⁹ Синтаксическая неоднозначность широко встречается в публицистических текстах, особенно в заголовках информационных сообщений, ср. вынесенное в заголовок предложение *У пострадавшего от нападения балетмейстера Филина улучшилось зрение* (<http://news-w.com/62851-18021/n/62851-8631/>), которое допускает три синтаксических прочтения.

В описываемом лингвистическом процессоре синтаксический анализ работает с ближайшим линейным контекстом с целью связать между собой соседние элементы. Начиная от левого конца предложения к его правому концу идет последовательная сборка токенов в синтаксические конструкции. Как правило, этот процесс включает несколько проходов. Это обусловлено, во-первых, тем, что синтаксические правила ранжированы нами по группам. Сначала в масштабе всего предложения происходит объединение токенов, связанных сильными локальными связями, затем — более слабыми и/или более удаленными и т. д. Во-вторых, каждый проход анализатора предполагает работу с окном вплоть до пяти элементов³⁰. Очевидно, что, если предложение состоит из шести и более токенов, границы этого окна будут поэтапно смещаться от левого к правому краю предложения (1–5, 2–6, 3–7 и т. д.). В-третьих, каждый раз в рамках одной и той же пятерки элементов может осуществляться поиск релевантных связей не только на максимальном удалении друг от друга (т. е. в рамках всего окна), но и на меньшем расстоянии: если не найдено правило для пяти элементов, далее осуществляется поиск подходящего правила для первых четырех элементов, затем первых трех и, наконец, первых двух элементов. В тот момент, когда какие-то элементы (на начальном этапе — токены) объединились в конструкцию, они, в свою очередь, становятся единым элементом, и окно тем самым «растягивается». Таким образом, при укрупнении элементов мы сближаем между собой сколь угодно удаленные токены и в итоге собираем все предложение³¹.

Результат конструктивно-синтаксического анализа представляет собой последовательность групп ПСДО: подлежащего, сказуемого, дополнений (прямого и косвенных) и обстоятельственных детерминантов, относящихся ко всему предложению в целом³². При же-

³⁰ «Элемент» здесь является понятием относительным: он может представлять собой либо одиночный токен, либо ранее собранную конструкцию.

³¹ Справедливости ради следует заметить, что для сборки целого предложения пяти элементов не всегда хватает — хотя бы потому, что в русском языке есть шестивалентный глагол *командировать*, не говоря уже о некотором числе пятивалентных. Если в предложении имеется вдобавок обстоятельственный (-е) детерминант (-ы), очевидно, что оно не поместится целиком в пятиэлементное окно. Поэтому при сборке целого предложения (и только в этом случае) используется более широкое окно.

³² Детерминант понимается здесь несколько иначе, чем у автора данного термина, ср. [Шведова 1990]. Основным критерием для нас является фактор отнесенности соответствующего слова или конструкции к предложению в целом,

лании указанные группы можно легко организовать в древовидную структуру. Очевидно, что групп дополнений, так же как и групп обстоятельств, в предложении может быть больше одной. Это же верно и в отношении подлежащего, если оно представлено рядом однородных членов.

Если предложение содержит однородные сказуемые, ему соответствует более одного ПСДО (по числу сказуемых). Так же обстоит дело со сложносочиненными предложениями и сложноподчиненными предложениями с расчлененной структурой. Сложноподчиненным предложениям с присловным придаточным приписывается одна структура ПСДО. Если предложение синтаксически неоднозначно, оно тоже имеет более одного ПСДО.

Каждая из групп ПСДО может быть реализована в тексте посредством многочисленных разновидностей, которые получают формализованное описание. Так, группа подлежащего может состоять из одного или нескольких однородных существительных, местоимений, собирательных числительных, инфинитивов, а также включать разнообразные сильные и слабые управляемые члены, в том числе:

- зависимые существительные, выражающие объектные или определительные значения, с предлогом или без (*дом отца, конь в яблках, влияние глобализации на климат Земли, дом у реки*);
- согласованное прилагательное (*красивый дом*), причастие (*разрушенный дом*), местоимение (*мой дом*) или порядковое числительное (*второй дом*);
- количественное или собирательное числительное (*два дома, оба дома, трое солдат*);
- существительное или именную группу в функции приложения (*река Москва, генерал от инфантерии Михаил Дмитриевич Скобелев*);
- определительный оборот (*белые от страха глаза*), в том числе причастный (*дом, покосившийся на один бок от старости*), в пре- или постпозиции;

а не к какому-то его отдельному члену. Вследствие этого в понятие детерминанта включаются не только собственно обстоятельственные обороты (за исключением обстоятельств образа действия, связанных более с предикатом, нежели с ситуацией в целом), но и ряд других разнородных феноменов — обращения, междометия, вводные слова и конструкции. Сюда же относятся модальные частицы, которые передают субъективное отношение автора к сообщаемому, а значит, «суперсегментны по своей сути» [Николаева 1985: 15]. Таким образом, «обстоятельственный детерминант» является для нас удобным «зонтичным» рабочим термином.

- определительное придаточное предложение (*дом, который я помнил с детства; тот, к кому я обратился с вопросом*);
- изъяснительное придаточное предложение с указательным словом и без (*слух (о том), что открылась выставка*);
- примыкающий инфинитив (*совет пойти учиться*), наречие (*взгляд исподлобья, дом напротив*), существительное или числительное (*дом номер восемь, дом восемь*).

Кроме того, возможны разнообразны комбинации перечисленных выше случаев (*два старых разрушенных дома, старый отцовский дом, который я помнил с детства* и т. д.).

Обратимся теперь к подробному рассмотрению процесса сборки предложения при помощи синтаксических правил. Правила осуществляют объединение элементов в некоторую результирующую конструкцию (в дальнейших манипуляциях она, в свою очередь, выступает в качестве элемента). При этом каждый элемент должен удовлетворять определенным условиям, которые фактически представляют собой значения признаков (графематических, морфологических, семантических и синтаксических)³³.

Допустим, что мы рассматриваем правила на два элемента. В этом случае речь может идти о конкретных значениях признаков для обоих членов правила (например, примыкание наречия с фиксированной позицией к главному слову, ср. *уха по-царски*), либо об одном конкретном значении, а другом относительном (так, род, число и падеж прилагательного (-ых) должны совпадать с соответствующими характеристиками определяемого существительного), либо, наконец, об относительных (взаимозависимых) значениях признаков для обоих элементов (например, в сочетании предлога с существительным требуется, чтобы падеж существительного совпадал с падежом, которым управляет предлог).

Результирующему объекту правило приписывает значения грамматических и семантических признаков. Так, словосочетанию

³³ То обстоятельство, что каждый токен несет на себе приписанные ему ранее значения не только морфологических, но и графематических признаков, является важным и способствующим правильной интерпретации предложения преимуществом описываемого лингвистического процессора. Если информацию, передаваемую знаками препинания, не принимать в расчет, может возникнуть ложная синтаксическая омонимия (классический пример *казнить нельзя помиловать* отнюдь не является единственным в своем роде). Подробнее о роли пунктуации в разрешении неоднозначности в системах автоматической обработки текста см. [Бердичевский, Иомдин 2007].

дубовый стол будет приписана следующая информация: «тип конструкции = существительное с согласованным определением, семантические признаки: базовый: предмет; мебель +, дерево +». Как мы видим, здесь произошло объединение значений признаков обоих слов. Однако так бывает не всегда: характеристика результирующего объекта может отличаться от характеристик членов конструкции. В качестве примера укажем на сочетания существительного с предлогом, «рождающие» новый смысл: если *стол* — это «часть речи = существительное; род = м; число = ед; падеж = им/вин», то *на стол, в стол* — это «тип конструкции = обстоятельство-места-куда»³⁴.

Синтаксические правила сгруппированы нами в соответствии с их рангом. Правила первого ранга, как было отмечено выше, — это правила объединения токенов, связанных наиболее тесными локальными связями. Правила второго ранга служат для объединения элементов (часть из которых могут представлять собой конструкции, собранные на предыдущем уровне), которые связаны менее сильными связями. Правила третьего ранга направлены в основном на присоединение обособленных оборотов и придаточных предложений к распространяемому члену (главного) предложения.

Разумеется, четкое распределение правил по уровням в известной мере идеализирует реальное положение дел: ведь для того чтобы создать жесткую последовательность их применения, необходимо исчислить возможные ситуации взаимного расположения токенов, учитывая весь набор их признаков, от графематики до семантики, но подобного описания нет, и едва ли оно возможно в принципе. Возьмем, к примеру, последовательность «существительное + существительное в родительном падеже». Если она стоит в конце предложения, можно утверждать, что мы имеем дело с генитивным оборотом, но когда она встречается в другой позиции, такой уверенности уже нет. Ср.: *На собрании выступил плотник совхоза* vs. *Плотник совхоза не мог покинуть*. Бывает так, что после выполнения правил

³⁴ Ср.: «семантический анализатор в процессе обработки исходного текста выполняет два основных действия: выбор правильного значения (предиката — формального аналога понятия) слова и построение из выбранных значений более сложного предиката. Это построение всегда носит характер функционального взаимодействия, в результате которого вычисляется новое значение, в частности — новый семантико-грамматический тип построенной конструкции» [Тузов 2003: 6].

более высокого ранга приходится возвращаться к правилам более низкого ранга.

Опишем теперь классические случаи, подпадающие под действие правил различных рангов. К правилам первого ранга относятся следующие:

- существительное + согласованное с ним одно или несколько полных прилагательных;
- предлог + существительное в соответствующей падежной форме³⁵;
- предлог + согласованное с существительным прилагательное + существительное в соответствующей падежной форме³⁶;
- частица *не* + следующее за ней знаменательное слово.

Здесь следует сделать оговорку. С целью сокращения числа правил связывания некоторые части речи и/или лексико-грамматические разряды могут объединяться в более крупные классы слов в соответствии с их синтаксическим поведением. Так, согласование с определяемым существительным по роду, числу и падежу, помимо прилагательных, наблюдается у причастий, порядковых числительных, некоторых типов местоимений. Все эти разряды слов в некоторых правилах совокупно считаются «прилагательными». С другой стороны, есть такие местоимения (так называемые местоимения-существительные (Русская грамматика 1980)), которые ведут себя в предложении подобно существительным. В формулировке некоторых правил термин «существительное» также понимается именно в таком расширенном смысле.

К правилам первого ранга относятся также те, что направлены на сборку аналитических форм глагола, а также глаголов с фазовыми и модальными операторами, прилагательных и наречий в сравнительной и превосходной степенях, составных числительных, Ф. И. О. в полном виде (в том числе с буквенными сокращениями типа *тов.*, *проф.*, *д-р*, *г-н* в препозиции), аппозитивных именных групп (*генерал Иванов, отличник Петя, газета «Известия»*). Таким образом, выполнение правил первого ранга дает нам то, что по ана-

³⁵ Информация о том, какого падежа требует тот или иной предлог, поступает из Семантико-синтаксического словаря.

³⁶ Данное правило фактически является суммой двух предыдущих; в качестве самостоятельного оно вводится для того, чтобы не упустить возможную интерпретацию токенов типа *больной* в определенных контекстах (пример см. ниже).

логии с графематическим анализом мы можем назвать «синтаксическими атомами».

Сборка токенов в ту или иную конструкцию может быть затруднена из-за специфики синтаксического контекста. Так, для вышеупомянутого правила объединения существительного и согласованного с ним прилагательного представляет сложность ситуация, в которой прилагательное «зажато» между двух существительных и формально согласовано с ними обоими. Она хорошо иллюстрируется известной строкой из стихотворения А. С. Пушкина *Роняет лес багряный свой убор*³⁷, где прилагательное *багряный* может относиться к любому из существительных *лес* и *убор*. Этот вид неоднозначности неразрешим в рамках конструктивно-синтаксического анализа предложения.

Другая пушкинская фраза *Друзья кровавой старины народной чаяли войны* (поэма «Полтава») также формально неоднозначна, ср.:

- 1) *друзья (кровавой старины)* — генитивная конструкция, где зависимое существительное имеет при себе согласованное определение;
народной (чаяли) войны — нарушенная проективность в сочетании сказуемого с группой прямого дополнения;
- 2) *друзья (кровавой старины народной) чаяли войны*.

Второй вариант интерпретации, однако, отклоняется в силу введенного нами ограничения, согласно которому все согласованные с существительным полные прилагательные должны располагаться либо слева от него, либо справа. Нахождение одного из них в препозиции, а другого в постпозиции возможно разве что в разговорной речи, причем, как правило, в конце предложения, ср.: *Купил карамельки и шоколадных конфет вкусных*.

Следует заметить, что автоматический анализ структуры пушкинских строк даже при условии сохранения только первого варианта все равно затруднен вследствие разрыва синтаксической конструкции «прилагательное + существительное» (*народной чаяли войны*).

³⁷ Хотя описываемый лингвистический процессор не предназначен для автоматического анализа поэтической речи (см. введение), в изложении мы используем отдельные примеры из стихотворных произведений в силу их широкой известности; сконструировать структурно идентичный прозаический пример не составляет труда.

Трактовка последнего примера демонстрирует важный аспект используемых нами правил, а именно их сугубо эмпирический характер, обусловленный внимательным наблюдением над языковым материалом. Собственно говоря, у нас не было другого пути — как отмечалось выше, готового перечня интерпретаций для всевозможных линейных комбинаций из слов различных частей речи просто не существует. В связи с этим приведем еще одно наблюдение, касающееся обработки сочетания «существительное + согласованное с ним полное прилагательное». Нами было установлено, что если между прилагательным и существительным «вклинивается» предложно-падежная форма, она является несогласованным определением к тому же существительному, ср. *черная в горошек блузка*. Иными словами, предложно-падежная форма в таком контексте не нарушает проективности — в отличие от глагола (*народной чаяли войны*). Последовательность обработки такого словосочетания следующая: сначала собирается синтаксическая группа *в горошек*, а затем она включается в конструкцию «согласованное определение + несогласованное определение + определяемое существительное»³⁸.

Грамматическая омонимия слов типа *больной, военный, пожарный, слепой* потребовала введения упомянутого выше правила «предлог + согласованное с существительным прилагательное + существительное в соответствующей падежной форме», которое обеспечивает успешный анализ предложений типа *К хрому мужику подошли двое*. Если бы мы располагали исключительно правилами на два элемента «предлог + существительное» и «прилагательное + существительное», то было бы ошибочно собрано словосочетание *к хрому* (*хромой* как существительное), а словоформа *мужику* оказалась бы «подвешенной», и предложение не было бы собрано. Наличие данного правила обеспечивает сборку всей именной группы *к хрому мужику*, которая в дальнейших операциях выступает в качестве единого элемента.

Данный пример также демонстрирует преимущество выбранного нами размера окна (подробнее об этом см. ниже). Если бы окно было размером в два элемента, правил на три элемента просто не существо-

³⁸ Вся эта процедура повторяется, когда подобные конструкции вложены одна в другую, ср.: *Он проглатывает слюну и смотрит на меня поверх очков **небольшими**, с нависшей над веками кожей глазами*» (В. Некрасов. В окопах Сталинграда).

вало бы, и нам пришлось бы обходиться вышеупомянутыми правилами на два элемента. Тогда также произошла бы ошибочная сборка словосочетания *к хромоту*, не позволяющая в итоге собрать предложение целиком. Окно же в пять элементов позволяет не только вместить именную группу целиком, но и увидеть следующий за ней глагол *погоди*, который, собственно говоря, и сигнализирует ее конец.

Грамматическая омонимия слов типа *больной, хромоты, слепой* может приводить к синтаксической неоднозначности, ср.: *Про больного доктора говорили, что он вряд ли поправится*. В целях ее сохранения параллельно применяются соответствующие правила на два и три элемента, что приводит к сборке следующих структурных вариантов предложения:

- 1) *про больного* (предлог + существительное) *доктора* (подлежащее) *говорили...*;
- 2) *про больного доктора* (предлог + прилагательное + существительное) *говорили...* (неопределенно-личное предложение).

Оба эти варианта оказываются допустимыми. Это может показаться парадоксальным, но сохранение присущей естественно-языковому тексту синтаксической омонимии также является важной задачей автоматического анализа текста, хотя сталкиваться с ней приходится существенно реже, чем с обратной проблемой — снятием неоднозначности.

Размер окна в пять элементов выбран в качестве оптимального сугубо эмпирически. Для языка со свободным порядком слов очевидна необходимость учета как левого, так и правого контекста, что в качестве минимума дает окно в три элемента. В принципе, можно было бы обойтись и этим размером, но во многих случаях он тормозит выявление синтаксических конструкций. Возьмем, к примеру, последовательность из трех токенов *красные и белые*. Можно предположить, что это однородные члены предложения. Но, во-первых, мы не можем определить, прилагательные это или существительные. Во-вторых, мы не знаем, весь ли это однородный ряд целиком или за рамками окна остались еще какие-то его члены (ср. *зеленые, красные и белые шарики*). Следовательно, мы не можем его сразу собрать. Заметим, что окна в пять элементов оказались бы достаточно, чтобы вместить весь этот однородный ряд целиком. При этом очевидно, что три однородных члена в предложении встречаются довольно часто — в отличие от пяти и более

членов. Наконец, вовсе не обязательно мы имеем здесь дело с однородными прилагательными, ср.: *кубики красные и белые шарик*и. В конечном итоге даже при окне в три элемента эту конструкцию удастся собрать, но скорость обработки существенно снизится. Еще одним фактором в пользу расширения окна являются частые нарушения проективности в русском языке. В то же время чрезмерное увеличение размера окна (до семи и более элементов) негативным образом сказывается на быстродействии анализатора. За счет многопроходности окно в пять элементов позволяет собрать все синтаксические группы (ПСДО) предложения³⁹.

Как уже говорилось, поиск подходящего правила в рамках текущего окна всегда начинается для максимально возможного числа элементов. Иными словами, сначала ищется правило на пять элементов (текущее + следующие четыре справа от него). Такое правило у нас одно-единственное, а именно «предлог + три полных прилагательных, разделенных пробелами (без знаков препинания) и согласованных с определяемым существительным + существительное в соответствующей падежной форме», ср. *Мы подошли [к большому белому деревянному дому]*. Когда соответствующая конструкция собрана, она становится единым элементом, который в дальнейших манипуляциях выступает в качестве цельного блока. Если подобной ситуации в тексте не обнаружено, окно сужается на один элемент и ищутся правила, которые могут быть применимы к текущей последовательности из четырех элементов. Прежде всего, это правило аналогично тому, что приведено выше для пяти элементов, с тем исключением, что вместо трех неоднородных прилагательных в нем фигурируют два. Соответственно, для трех элементов существует правило «предлог + прилагательное + существительное», для двух элементов — «предлог + существительное» (о них уже шла речь выше). Разумеется, имеются и другие правила для сборки конструк-

³⁹ При условии правильно построенного предложения. Как бы ни было условно и растяжимо это понятие, все же оно имеет границы, и их нарушение ощущается носителями языка. Время от времени в печати появляются предложения, анализ которых представляет сложность даже для человека, не говоря уже о компьютере, ср.: *Как сообщили в ГУ МВД РФ по СПб и ЛО, минувшей ночью, около 00.10, у дома № 2 по Торфяной дороге неустановленный автомобиль совершил наезд на переходившего проезжую часть по пешеходному переходу полицейского управления вневедомственной охраны* (петербургская интернет-газета «Фонтанка.ру» 18.11.2012; газета «Метро» 18.11.2012).

ций из двух, трех и четырех элементов — некоторые из них были упомянуты выше.

Прежде чем рассмотреть пример сборки предложения посредством правил первого ранга, следует заметить, что, помимо «ранжированных» правил, есть еще особые правила, направленные на проверку полноты предложения, т. е. на их соответствие определенному структурному шаблону. Они выполняются перед запуском каждого следующего ранга правил, в том числе и в самом начале анализа, перед правилами первого ранга. Если структурная схема предложения соответствует тому или иному шаблону, применение правил сборки вообще не требуется.

Так, предложение *Мама мыла раму* соответствует шаблону: «существительное в именительном падеже + переходный глагол в личной форме, согласованный с существительным + существительное в винительном падеже без предлога», — что позволяет сразу собрать его, не привлекая правила каких бы то ни было рангов. Кроме того, это уберегает нас от ошибочной сборки *мама мыла* как «существительное в именительном падеже + существительное в родительном падеже» (правило второго ранга, см. ниже) и, следовательно, «лишней» омонимии. Шаблон допускает перестановку элементов, ср.: *Мама раму мыла* и т. д.

В структурно схожем примере *Директор школы не любил* проверка полноты предложения первоначально не дает результата. После применения правила первого ранга для объединения частицы *не* с глаголом *любил* в единый элемент проверка осуществляется вновь и на этот раз приводит к сборке предложения в соответствии с тем же шаблоном, что был использован в случае *Мама мыла раму*⁴⁰. Примечательно, что автоматический анализ здесь не допускает ошибки, обыкновенно совершаемой носителем языка при обработке подобных предложений (которые в англоязычной лингвистической литературе принято называть *garden-path sentences*).

Шаблон, актуализированный в рассмотренных примерах, на самом деле является частным случаем общего рабочего принципа, связанного с проверкой заполнения актантных позиций глагола. Если предложение состоит исключительно из глагола и именных

⁴⁰ Наличие отрицания при глаголе существенно в том смысле, что допускает употребление дополнения как в форме беспредложного винительного, так и родительного падежей. В данном случае следствием является неоднозначность формы *школы*: либо родительный падеж единственного числа, либо винительный падеж множественного числа.

групп, морфологическая форма и семантическое содержание которых полностью соответствует условиям реализации какого-либо из глагольных ССВ, то предложение собирается сразу, ср. *Петя забил гвоздь в стену молотком*. Такая ситуация, правда, встречается нечасто.

Более распространенными являются следующие ситуации. Первая: число именных групп в тексте меньше, чем число актантов, указанных в условиях реализации, ср. *Петя забил гвоздь в стену*. Если морфологическое оформление и семантическое содержание наличных актантов совпадает с условиями реализации, предложение также считается полным и собирается сразу. Вторая: число именных групп в тексте больше, чем число заполненных условий реализации, ср. *Петя забил гвоздь в стену молотком + утром/на даче/для развлечения /....* Если «лишняя» именная группа представляет собой нераспространенный обстоятельственный детерминант, относящийся к предложению в целом, оно также собирается сразу. В том случае, когда мы имеем дело с обстоятельством образа действия, относящимся непосредственно к глаголу (ср. *Петя легко/с трудом/с первой попытки забил гвоздь в стену молотком*), требуются дополнительные шаги по сборке группы данного обстоятельства (если оно распространенное) и присоединению ее к сказуемому.

Проиллюстрируем теперь последовательный процесс сборки синтаксической структуры на примере предложения *Маша купила черную в белый горошек блузку*.

0 — не найдено правило сборки полного предложения.

1.1. [*Маша купила черную в белый*] *горошек блузку* — правило не найдено.

1.2. [*Маша купила черную в*] *белый горошек блузку* — правило не найдено.

1.3. [*Маша купила черную*] *в белый горошек блузку* — правило не найдено.

1.4. [*Маша купила*] *черную в белый горошек блузку* — правило не найдено.

2.1. *Маша [купила черную в белый горошек]* *блузку* — правило не найдено.

2.2. *Маша [купила черную в белый]* *горошек блузку* — правило не найдено.

2.3. *Маша [купила черную в]* *белый горошек блузку* — правило не найдено.

2.4. *Маша [купила черную] в белый горошек блузку* — правило не найдено.

3.1. *Маша купила [черную в белый горошек блузку]* — правило не найдено.

3.2. *Маша купила [черную в белый горошек] блузку* — правило не найдено.

3.3. *Маша купила [черную в белый] горошек блузку* — правило не найдено.

3.4. *Маша купила [черную в] белый горошек блузку* — правило не найдено.

4.1. *Маша купила черную [в белый горошек блузку]* — правило не найдено.

4.2. *Маша купила черную [в белый горошек] блузку* — правило найдено. Собрана предложно-падежная группа.

5.1. [*Маша купила черную* (предложно-падежная группа) *блузку*] — правило не найдено.

.....
6.1. *Маша [купила черную* (предложно-падежная группа) *блузку]* — правило не найдено.

.....
7.1. *Маша купила [черную* (предложно-падежная группа) *блузку]* — правило найдено! Собрана группа существительного с согласованным и несогласованным определениями.

8. [*Маша купила* («существительное с определениями»)] — найдено правило сборки полного предложения. Предложение собрано: существительное в именительном падеже + переходный глагол в личной форме, согласованный с существительным + существительное в винительном падеже без предлога.

Применение правил первого ранга повторяется для предложения до тех пор, пока срабатывает хоть одно правило. Если не найдено ни одного правила, то работа с правилами первого ранга заканчивается.

* * *

Правила второго ранга имеют своей целью сборку следующих конструкций:

- однородные члены предложения (*синие, зеленые, красные и белые шарики; отличник Петя, хромой мужик и генерал Иванов*⁴¹);

⁴¹ После применения правил первого ранга каждый член однородного ряда представляет собой целостный элемент.

- существительное + существительное в родительном падеже (генитивная конструкция);
- количественное числительное + существительное в соответствующем падеже;
- даты, представленные нецифровым шаблоном (*пятнадцатому мая, 8 марта, 28-го октября, 13 апреля 2012 г.*⁴²).
- число, записанное буквами или цифрами (с буквенным окончанием или без) + буквенное сокращение + существительное в родительном падеже (*два кг яблок, 2 кг яблок, 115-й км трассы*).
- наречие + главное слово.

Обрисует вкратце проблемы, которые могут возникнуть при применении правил сборки данных конструкций.

Применительно к сочинительным конструкциям следует прежде всего отметить возможность несовпадения морфологических характеристик у однородных членов и даже их функциональной разнотипности, что очевидным образом затрудняет автоматический анализ, ср.:

Купил чаю и коробку спичек;

Впереди стоял комендант, старик бодрый и высокого роста (А. С. Пушкин. Капитанская дочка);

Прощай все и навсегда (И. С. Тургенев. Дневник лишнего человека);

В Москве и без дела он уже жить не может (С. А. Толстая);

Уехал, но ненадолго;

Уехал, и надолго.

Весьма заманчивой выглядит идея создания исчерпывающего списка сочетаний, допустимых для той или иной синтаксической функции, которая, как утверждает В.З.Санников [1989: 46], реализована им в более ранней работе [Санников 1981]. Вместе с тем широкая вариативность грамматического оформления однородных членов, даже если они выступают в одной функции (свидетельством чего могут служить как примеры из [Санников 1989], так и материал НКРЯ), вызывают сомнение в осуществимости данной задачи в пол-

⁴² Даты, записанные цифрами в соответствии с шаблонами (типа 13.02.2012), выявляются, проверяются и атрибутируются на предшествующих этапах анализа.

ном объеме, не говоря уже о случаях функциональной разнотипности.

Другая хорошо известная проблема — согласование сказуемого с сочиненными подлежащими — применительно к задаче автоматической обработки текста была внимательно рассмотрена В.З. Санниковым [1989] и Л.Л. Иомдиным [1990]. Поскольку нас интересует только анализ текста, но не его синтез, вопрос связан не с выбором нужной формы глагола (*Остался/остались девочка и мальчик*), а с вариативностью в числе сказуемого. В соответствующих синтаксических правилах приходится допускать различные варианты согласования сказуемого с сочиненными подлежащими по числу, роду и лицу, ср.:

Пришел/пришли Миша и Петя;

Почти миллион человек работал/работало/работали в сельском хозяйстве;

*Кто и что мешает (*мешают) тебе работать?;*

*Хорошее и плохое смешалось (*смешались) в памяти;*

То ли чурка, то ли бочка проплывает по реке (А. Твардовский. Василий Теркин);

*Приехала (*приехали) не только Маша, но и Петя;*

Швед, русский колет, рубит, режет (А.С. Пушкин. Полтава);

*Швед и русский поскакали (*поскакал) к реке.*

В случае употребления глагола в единственном числе дополнительная сложность связана с определением того, кореферентны ли однородные подлежащие, ср.:

Прямо у крыльца росла большая сосна и дуб (два дерева);

Выступил известный музыкант и композитор (один и тот же человек).

В предложениях наподобие последнего встает также вопрос о «сфере действия» зависимых членов. Характеризует ли согласованное определение *замечательные* в примере *Здесь работают замечательные писатели и врачи* представителей обеих профессий? К чему в предложении *На пороге показались Коля и Вася с цветами* в руках относится несогласованное определение в постпозиции —

к ближайшему подлежащему или к обоим однородным членам? В записях на формальных языках в подобных случаях используются скобки, позволяющие снимать неоднозначность, но в ЕЯ она неизбежна.

Приходится также принимать во внимание возможность семантического, а не грамматического согласования сказуемого с подлежащим по числу, ср. *Большинство родителей не смогли прийти на собрание; На тропинке показались отец с сыном; Новый премьер-министр обратилась к собравшимся с речью.*

Еще одна трудность состоит в том, что сочиненные именные группы (так же, как и множественные именные группы) могут интерпретироваться и в объединительном, и разделительном смысле. Так, предложение *Петя и Коля получили грант в размере трех тысяч долларов* может осмысляться в двух значениях: ‘три тысячи всего’ и ‘три тысячи каждому’ [Богуславский 1996: 140–141]. Если предложение содержит несколько сочиненных групп, ситуация еще более запутывается, ср. *Вчера и позавчера Аня и Володя видели Лену и Алешу в кафе и на дискотеке*⁴³. Очевидно, что автоматический анализ, построенный на комбинаторном исчислении возможных вариантов, не способен выявить ту наиболее очевидную интерпретацию, которая сразу приходит на ум человеку.

Более того, вследствие недостаточной эксплицитности и точности языковых выражений логическое исчисление способно порождать ложные варианты. Например, из предложения *Коля и Петя купили в киоске французские, английские, немецкие газеты и журналы* вовсе не следует, что каждый из субъектов купил весь спектр печатной продукции (шесть видов). Предложение содержит утверждения о группах объектов, а не о каждом индивидуально. Однако дальнейшее развитие повествования может касаться любого из данных объектов — Коли, Пети, английских газет, немецких журналов и т. д. Поэтому на следующем этапе анализа (коммуникативно-синтаксическом) мы выделяем все восемь объектов по отдельности (Коля,

⁴³ Схожая проблема возникает и при интерпретации сочинительной связи между придаточными предложениями. Так, предложение *Давайте встретимся с избирателями, которые живут в этом районе и поддерживают «Яблоко»* предполагает конъюнкцию предикатов (*живут и поддерживают*). В отличие от него, предложение *При описании наблюдений следует избегать категоричности выводов, если число наблюдений недостаточно и если в наблюдениях имеются неясности или сомнения в точности наблюдений* допускает и конъюнкцию, и дизъюнкцию пропозиций, выраженных условными придаточными [Тискин 2016].

Петя, французские газеты, английские газеты, немецкие газеты, французские журналы, английские журналы, немецкие журналы), причем каждый вид газет и журналов связывается отношением покупки как с Колей, так и с Петей. Те или иные связи могут быть отменены при наличии опровергающей информации в последующих предложениях.

Что касается второго из перечисленных правил, связанного со сборкой генитивной конструкции, проблема заключается прежде всего в определении хозяина беспредложного генитива: ведь он может управляться как глаголом, способным присоединять соответствующий актант, так и другим существительным. Встречаются случаи синтаксической омонимии, разрешить которую может лишь учет значений семантических признаков всех «претендентов». Иными словами, единого правила, связанного с обработкой беспредложного генитива, не существует⁴⁴. Эта ситуация, вообще говоря, вполне типична. То, что мы для краткости изложения формулируем в качестве правила сборки конструкции, на деле часто представляет собой совокупность инструкций, касающихся действий анализатора в частных случаях того или иного расположения токенов с приписанными им значениями признаков.

Если все же установлено, что мы имеем дело с генитивной конструкцией, может возникнуть еще один источник синтаксической неоднозначности, обусловленный возможностью двоякого прочте-

⁴⁴ Неоднозначность, связанная с конкуренцией между глаголом и существительным за право управлять именной группой, естественно, не ограничивается формой беспредложного родительного. Так, в предложении *Но нет уже в них того азарта и самоуверенности, и все реже и реже сбрасывают они на наши головы тучи листовок с призывами сдаться и бросить надежды на идущего с севера Жукова* (В. Некрасов. В окопах Сталинграда) предложно-падежная группа «на + S_{вин.}» с формальной точки зрения может быть актантом как глагола *бросить*, так и существительного *надежды*. Однако для того чтобы она относилась к глаголу, необходимо, чтобы объектная валентность, выраженная прямым дополнением, была заполнена существительным с предметным значением (*бросить ключи на стол*). С другой стороны, предикатное существительное *надежда* может управлять практически любым существительным в данной позиции. В данном случае информация о сочетаемостных свойствах лексических единиц из Семантико-синтаксического словаря позволяет снять синтаксическую омонимию. Иначе обстоит дело со следующим заголовком информационного сообщения: *Мужчина прыгнул с моста на остров Русский во Владивостоке* (<http://rusnovosti.ru/news/255776/> (дата обращения: 24.03.2018)), — который неоднозначен в принципе.

ния — субъектного и объектного, ср. *встреча друзей, портрет художника* и т. д.

Дальнейшая работа по анализу генитивной конструкции касается ее семантической интерпретации. Как известно, конструкция с родительным падежом способна выражать широчайший спектр значений, в том числе метафорических [Кнорина 1990; Михеев 2000]. Семантический анализ генитивной конструкции в каждом конкретном случае является весьма сложной задачей, сопоставимой разве что с определением значения относительных прилагательных в сочетании с различными существительными.

Некоторых комментариев требует правило присоединения наречия к главному слову. Как известно, наречия могут примыкать к глаголу (включая причастие и деепричастие), прилагательному, другому наречию и существительному, так что, в принципе, может возникать ситуация конкуренции между его потенциальными «хозяевами». Однако, во-первых, сама суть синтаксического отношения примыкания предполагает контактное расположение слов в словосочетании. Во-вторых, для каждого наречия существует предпочтительная (а иногда жестко фиксированная) позиция по отношению к своему синтаксическому хозяину. И если в устной речи нарушение предпочтительной позиции «компенсируется» использованием просодических средств, то в письменном тексте это невозможно, а следовательно, она соблюдается практически всегда. К примеру, в спонтанной речи за счет определенного интонирования и паузации наречие *очень* в предложении *Девочка красивая очень смутилась и покраснела* может быть синтаксически прикреплено к прилагательному *красивая*, но в тексте однозначно воспринимается как относящееся к глаголу *смутилась*.

* * *

Правила третьего ранга имеют своей целью сборку обособленных членов предложения и их прикрепление к главному слову. Это касается одиночных причастий и деепричастий, причастных и деепричастных оборотов, обстоятельственных оборотов, выраженных именем существительным с предлогом, согласованных и несогласованных определений, приложений, пояснительных и уточняющих оборотов. Разработаны определенные правила их прикрепления к тому или иному члену результирующей синтаксической структуры предложения. Так, обособленные причастия, причастные и другие адъективные обороты могут прикрепляться к группе подлежащего

(П) или группе дополнения (Д). Некоторые сложности здесь могут возникнуть с обработкой несогласованных определений, в силу того что они имеют слабую связь с определяемым существительным, а синтаксическая интерпретация слабоуправляемых членов предложения, как будет показано ниже, не имеет универсального решения.

Обособленные деепричастия и деепричастные обороты преобразуются в самостоятельную структуру ПСДО: сказуемому соответствует деепричастие, а подлежащее дублируется из ПСДО главной пропозиции. При этом, правда, неэксплицированной остается обстоятельственная семантика деепричастий, или (в проекции на наш формат синтаксического представления) содержательная связь между ПСДО с личным глаголом и ПСДО с деепричастием. (Впрочем, это часть более общей и многосторонней проблемы о связях между пропозициями в тексте, затрагиваемой в следующей главе.) Что же касается обособленных субстантивных обстоятельственных оборотов (*во избежание падения, по причине отсутствия* и т. д.), они преобразуются в группу обстоятельственных детерминантов (О)⁴⁵.

По-видимому, не существует единого подхода к синтаксическому анализу обособленных приложений. В некоторых контекстах такое приложение содержательно относится ко всему предложению в целом [Камынина 1983: 80–81], и, следовательно, согласно принятому нами подходу, его следует квалифицировать как обстоятельственный детерминант (О). Так, в примере **Человек молчаливый и очень точный**, он избегал скороспелых выводов и преждевременных обобщений обособленный оборот не столько характеризует лицо, сколько обозначает причину того положения дел, которое описывается главной пропозицией. Однако в предложении *Через полчаса явился уездный лекарь, человек небольшого роста, худенький и черноволосый* (И. Тургенев. Уездный лекарь)⁴⁶ структурно схожее приложение фактически служит определением к слову *лекарь*, приписывая данному лицу свойство 'невысокий'. Информация, передаваемая здесь обособленным оборотом, должна быть прикреплена к группе подлежащего (П), но формальные способы дифференциации рассмотренных двух случаев пока не ясны (предположительно начальная позиция оборота в предложении марки-

⁴⁵ Необособленное одиночное деепричастие, примыкающее справа к глаголу, входит в состав группы сказуемого (С).

⁴⁶ Примеры заимствованы из книги: [Учебно-методический комплекс... 2008: 161].

рует обстоятельственное прочтение, но эта гипотеза нуждается в дополнительной проверке).

Кроме того, правила третьего ранга осуществляют прикрепление придаточных предложений к опорному слову в главном предложении (для сложноподчиненных предложений с нерасчлененной структурой). Наиболее часто встречаются следующие два вида придаточных:

- 1) изъяснительные придаточные предложения с союзом *что*, заполняющие актантную позицию у предикатной лексики, обозначающей речевую и ментальную деятельность, эмоциональные состояния. Их прикрепление к опорному слову осуществляется с опорой на описание соответствующих слов в Семантико-синтаксическом словаре, где указаны возможности реализации объектной валентности не только посредством именных групп, но и сентенциальных актантов, вводимых различными союзами, а также прямой речи в кавычках;
- 2) определительные придаточные предложения, вводимые союзным словом *который*.

Указанные случаи достаточно легко распознаются и обрабатываются. Применительно ко второму типу, правда, следует упомянуть о ситуации, когда определительному придаточному предшествует не одиночное существительное в главном предложении, а генитивная конструкция из двух и более членов. Если существительные имеют один и тот же грамматический род, присоединение придаточного в принципе возможно к любому из них, ср.: *Директор института, который в прошлом году был награжден орденом Трудового Красного Знамени...* Нужно заметить, что данный пример является неоднозначным и для носителя русского языка⁴⁷ — в отличие от таких примеров, как:

Директор института, который расположен в городе...

Директор института, который еще недавно болел...

Однако в рамках конструктивно-синтаксического анализа эти примеры будут также интерпретироваться как неоднозначные,

⁴⁷ О психологических факторах, влияющих на интерпретацию определительной связи в подобных случаях, см. [Федорова, Янович 2004; Драгой 2006; Юдина 2006].

поскольку алгоритм не включает проверку на семантическое согласование слов *директор и болел, институт и болел* и т.д. Соответственно, будут построены различные варианты синтаксической структуры, присоединяющие придаточное определительное к одному и другому существительному. Выбор того из них, который является правильным, возможен лишь на следующем этапе анализа.

Обобщая ситуацию, связанную с автоматической обработкой сложноподчиненных предложений, можно сказать, что она происходит достаточно благополучно, если союз (или союзное слово) способен присоединять придаточное лишь какого-то одного определенного типа. Когда же связующий элемент характеризуется неоднозначностью (ср. *какой, кто, что, когда* и др.), правила сборки усложняются. Так, союзные слова *какой, кто, что* могут присоединять к главному предложению и изъяснительные, и определительные придаточные. Для дифференциации этих случаев приходится брать во внимание тип опорного слова (ср. *Я спросил, кто придет.* vs. *Тот, кто придет,...; Я спросил у него, какие книги нужны* vs. *Я собрал все книги, какие нужны*). Если слово нужного типа в главном предложении отсутствует, сборка осуществляется «по дефолту».

Приведем еще один пример. Токен *когда* в расчлененных сложноподчиненных предложениях выступает в качестве союза, присоединяющего придаточное времени, а в нерасчлененных служит для прикрепления определительных или изъяснительных придаточных, ср. *Когда стемнело, все легли спать* vs. *Наступает пора, когда рано светает* vs. *Мне нравится, когда задают много вопросов*. Идентификация этих трех вариантов также базируется на поиске в главном предложении лексических единиц определенных семантических групп. Так, изъяснительное придаточное выявляется по наличию глагола эмоционального состояния *нравиться*. Определительное придаточное удастся опознать благодаря существительному со значением временного промежутка (*пора*). Что же касается придаточных времени, они не имеют специфических показателей и идентифицируются по остаточному принципу. Поэтому если в главном предложении не обнаружен ни один из вышеуказанных маркеров, придаточному приписывается синтаксическое значение 'придаточное времени'.

Существуют также специальные правила третьего ранга, направленные на сборку сложных обстоятельственных детерминантов, чьи составляющие могут быть или не быть разделены запятой, напри-

мер: на Невском пр., в д. 35, в кв. 19; в Москве на Красной площади; в прошлом году весной; в пятницу вечером.

* * *

Схематические шаблоны предложений и ранжированные правила призваны обеспечить адекватный автоматический анализ наиболее частотных, центральных синтаксических структур в русском языке, которые имеют стандартную, регулярную интерпретацию (пусть и не обязательно однозначную). Следует признать, что многие факты языка неизбежно остаются за их рамками. Прежде всего, это разнообразные периферийные феномены, перечислять которые нет ни возможности, ни необходимости в силу их окказиональности и особенности. Но между этими двумя крайностями — выраженным центром и расплывчатой периферией — расположен целый ряд достаточно регулярных структур, которые легко опознаются и выделяются в тексте, но не имеют стандартной семантико-синтаксической интерпретации. Их успешная автоматическая обработка в каждом отдельном случае требует учета разнообразных контекстуальных факторов, причем конкретный набор этих факторов и характер обусловленности нуждается в тщательной проработке. Это работа, требующая больших затрат сил и времени.

Возьмем в качестве иллюстрации феномен так называемого слабого управления [Пешковский 1956], или падежного примыкания (Русская грамматика 1980)⁴⁸. Рассмотрим грамматически тождественные предложно-падежные группы *пить чай с женой* / *пить чай с вареньем* / *пить чай с удовольствием* и покажем, что они должны получить различную семантико-синтаксическую интерпретацию.

В выражениях типа *пить чай с женой*, *гулять с собакой*, *идет дождь со снегом*, предложно-падежная группа *с + N_{тв.}* выражает идею совместности действия. Такое значение приписывается данной конструкции тогда, когда оба существительных характеризуются однотипными значениями базового семантического признака (в наших примерах это 'человек' и шире — 'живое существо'; 'природная сила'). В этом случае предложно-падежная группа должна быть поднята в ранг однородного подлежащего, ср.: *я и жена пьем чай*, *я и собака гуляем*, *идут дождь и снег*. Разумеется, данное преобразование

⁴⁸ При разработке семантической интерпретации предложно-падежных групп, помимо толковых словарей русского языка, использовался Синтаксический словарь Г. А. Золотовой (Золотова 1988).

приводит к некоторому семантическому сдвигу, но с точки зрения информационной структуры текста им можно пренебречь. Схожая ситуация наблюдается в выражениях *пригласить профессора с его учениками, положить в пенал ручки с карандашами*, где рассматриваемая конструкция должна интерпретироваться как однородное дополнение, ср.: *пригласить профессора и его учеников, положить в пенал ручки и карандаши*.

Словосочетание *пить чай с вареньем* по своей сути является неоднозначным: то ли варенье находится в чае, то ли субъект попеременно отхлебывает чай из чашки и зачерпывает из вазочки (банки и т.п.) варенье. Первая интерпретация довольно типична для слов с семантическим признаком «еда, напитки = +», ср. *каша с маслом, суп с белыми грибами, макароны с сыром* и т.д. Обоснованность второго прочтения обусловлена однородностью объектов (см. выше). В ходе конструктивно-синтаксического анализа должны быть порождены обе версии.

Наконец, в выражении *пить чай с удовольствием* предлог сочетается с абстрактным существительным, а глаголу по базовому семантическому признаку приписано значение ‘действие’, следовательно, данная синтаксическая группа представляет собой обстоятельство образа действия.

Вообще, сочетание предлога *с* с творительным падежом, как известно, принадлежит к числу примеров, при помощи которых А.М. Пешковский [1956: 285] иллюстрировал введенное им понятие слабого управления. Рассмотренные случаи отнюдь не исчерпывают возможные реализации данной конструкции, ср. *бутылка с молоком, девочка с косичками, мыть с мылом, выехать с рассветом* и т.д. Тем не менее их достаточно для того, чтобы показать невозможность выработки единой процедуры анализа.

Слабое управление встречается довольно часто. По-видимому, практически любая предложно-падежная группа может в том или ином контексте реализовывать сугубо грамматическую зависимость, не обусловленную семантикой главного слова. Бывает, что несколько словосочетаний находятся в отношении слабого управления по отношению к одному и тому же глаголу или существительному, ср. название известного детективного романа С. Жапризо «Дама в очках с ружьем в автомобиле».

Отсутствие непосредственной связи с семантикой главного слова затрудняет синтаксический анализ подобных выражений, поскольку их появление в тексте не предсказывается условиями реализации,

описанными в семантико-синтаксическом словаре. В ряде случаев они повышают индекс омонимии предложения за счет неоднозначности синтаксического хозяина либо из-за возможности данного словосочетания выражать как семантическую валентность, так и чисто грамматическую зависимость.

Феномен слабого управления в лингвистике изучен недостаточно, соответствующие случаи весьма многообразны, требуют индивидуального подхода и тонкого анализа. Объем работы здесь поистине огромен, и мы ни в коей мере не претендуем на исчерпывающее описание.

Глава 4

КОММУНИКАТИВНО-СИНТАКСИЧЕСКИЙ АНАЛИЗ

ТЕОРЕТИЧЕСКИЕ ПРЕДПОСЫЛКИ

Коммуникативно-синтаксический анализ — заключительный этап работы лингвистического процессора, на котором происходит преобразование информации, представленной в виде ПСДО отдельных предложений, в метаязыковую структуру данных.

Неверно было бы думать, что данный этап знаменует переход от синтаксиса предложения к синтаксису текста («макросинтаксису» [Богданов 1977] или «гиперсинтаксису» [Откупщикова 1982]). Противопоставление конструктивного и коммуникативного синтаксиса лишь отчасти соотносится с понятиями предложения и текста соответственно: так, вопросы, связанные с разрешением анафоры, когезией и когеренцией, как правило, обсуждаются применительно к тексту; вместе с тем они могут возникнуть и на уровне предложения. С другой стороны, самые известные коммуникативные значения — тема и рема — выделяются в рамках предложения, однако очевидно, что «на коммуникативную перспективу очередного предложения могут оказывать влияние весьма обширные отрезки предыдущего текста» [Богданов 1977: 135].

К тому же понятие «предложение» традиционно определяется графическими признаками и потому является весьма условным. Так, сложные предложения могут без потери смысла быть разбиты на простые, парцеллированная конструкция может находиться внутри предложения либо быть оформлена последовательностью графически отдельных предложений и т. д. В связи с этим классические задачи лингвистики текста, связанные с выявлением локальных связей между соседними предложениями и выделением таких промежуточных единиц, как сложное синтаксическое целое [Поспелов 1948; Солганик 1973] и сверхфразовое единство [Булаховский 1952: 392–393; Гальперин 1981; Москальская 1981], не являются для нас приоритетными.

Специфика коммуникативно-синтаксического анализа обусловлена не его объектом, а содержательной направленностью. Его ко-

нечная цель — выявление информационной структуры текста на ЕЯ. Ограничения на тип текста сведены к минимуму — в качестве материала могут использоваться любые правильно построенные письменные тексты публицистического, научного, официально-делового стиля, а также художественной прозы. Эта ситуация кардинальным образом отличается от проектов автоматической обработки текстов с заранее известной «денотатной структурой» (ср. [Новиков 1983]). Заметим, что и популярные ныне системы извлечения частичной информации (*information extraction*) также ограничиваются поиском в тексте информации, касающейся заранее известных, заданных объектов и событий¹. Для этого сначала формируются списки соответствующих лексических единиц, затем для каждой единицы создается обобщенный образец ее поведения с точки зрения сочетаемости с другими лексемами (своего рода поисковое предписание). Поиск осуществляется на морфологически размеченных (тэгированных) корпусах текстов. Синтаксический компонент в таких системах, как правило, вовсе отсутствует, а семантика ограничивается заранее заданными списками и шаблонами [Леонтьева 2006: 185–189].

Наша постановка задачи принципиально иная — лингвистический процессор не знает, с каким текстом он работает (стиль, жанр) и о чем в этом тексте идет речь. Он должен уметь преобразовывать результат конструктивно-синтаксического анализа (ПСДО) отдельных предложений в новый формат — информационную структуру текста. Очевидно, что здесь мы вступаем в малоизученную область. Наши разработки в этом направлении носят отчетливо выраженный поисковый характер и не претендуют на решение всего спектра проблем, связанных с анализом текста как единого целого.

На данном этапе мы не можем довольствоваться исключительно сведениями об устройстве языковой системы. Ведь текст — это не очередной уровень языка, расположенный над предложением (как первоначально предполагалось в рамках структуралистской парадигмы), а феномен совершенно иной природы [Бенвенист 1974: 139]. Его организация регулируется не только лингвистическими, но и тематическими, стилистическими, функциональными факторами².

¹ Как правило, дело касается автоматической рубрикации текстов СМИ (ср., например, [Добров 2014]).

² Как писал Роман Якобсон, «в правилах сочетаемости лингвистических единиц существует ведущая вверх лестница свободы. В комбинировании различных признаков фонем свобода отдельного говорящего равна нулю: языковой код уже установил все возможности, которые могут быть реализованы в данном

Следовательно, для анализа текста требуется использовать опыт, накопленный в смежных дисциплинах, прежде всего когнитивных и дискурсивных исследованиях.

Дискурсивные исследования способны пролить свет на особенности структурной организации текстов различных типов. Многообразие речевых жанров и разная степень их структурированности препятствуют обстоятельному изучению этого предмета. Вместе с тем литература по анализу дискурса содержит полезные сведения, касающиеся закономерностей строения отдельных жанров и типов текстов, которые, в частности, могут помочь в интерпретации темпоральной структуры описываемых событий (см. ниже).

Что касается когнитивных исследований, они скорее определяют общую теоретическую платформу наших рассуждений, чем предоставляют конкретные сведения по тому или иному вопросу. Именно здесь, на этапе коммуникативно-синтаксического анализа, становятся актуальными центральные проблемы когнитивной науки, а именно: что люди знают о себе и о мире, откуда они знают то, что знают, как организовано это знание и как оно активируется (пускается в ход) [Gibbs 1996: 29, 40]. Эти вопросы в концепциях различных ученых-когнитивистов получают несколько разное преломление, ср. концептуальную метафору Дж. Лакоффа и М. Джонсона, аспекты образности Р. Лангакера, проблему категоризации, ментальные пространства Ж. Фоконье, оппозицию фигуры и фона и другие схематические системы у Л. Талми и др. (обзор данных концепций см., например, в [Скребцова 2011]). Тем не менее все когнитивные теории языка так или иначе сосредоточены на проблеме так называемого ментального конструирования, осуществляемого говорящим (при переходе от мысли к речи) и слушающим (при обратном переходе от речи к мысли). С точки зрения анализа текста особенно важна позиция слушающего; в то же время не следует пренебрегать и фигурой говорящего, который из доступного набора языковых средств по каким-то причинам выбирает именно те, что представлены в тексте. Последний аспект связан с такой важной категорией текста, как точ-

языке. Свобода соположения фонем в словах очень ограничена, она проявляется лишь в маргинальной ситуации словотворчества. В конструировании предложений и слов ограничений, налагаемых на говорящего, уже значительно меньше. Наконец, при создании текста из предложений действие ограничивающих синтаксических правил прекращается, и свобода каждого отдельного говорящего достигает максимума, особенно если он не должен придерживаться множества языковых клише» (цит. по: [Пеше 1999: 312]).

ка зрения, или перспектива (ср. [Успенский 1970; Шмид 2003: 109–144]), которая, в частности, влияет на интерпретацию дейктических выражений и обуславливает оценочные значения в тексте.

На первый взгляд это может показаться странным, но на данном этапе работы нам близок (если отвлечься от понятийного и терминологического аппарата когнитивизма) тот ракурс рассмотрения языка, его средств и механизмов, который принят в книге классика когнитивной лингвистики Ж. Фоконье [Fauconnier 1985]. Автор строит оригинальную теорию ментальных пространств, призванную описать динамику восприятия дискурса человеком. По Фоконье, этот процесс предполагает непрерывное конструирование мысленных конструкторов: создание и обновление ментальных пространств, введение новых элементов, добавление связей и т.д. При этом внимание исследователя сосредоточено на вопросе о том, какие типы информации, касающейся различных аспектов ментального конструирования, передаются теми или иными языковыми средствами. Не претендуя на исчерпывающее описание, автор отмечает достаточно очевидные случаи (в проекции на английский и французский языки). Например, неопределенный артикль несет информацию о создании нового элемента, грамматические значения времени и наклонения глагола сообщают о статусе ментального пространства, степени его выделенности и актуализованности, анафорические выражения отсылают к уже имеющимся элементам и т.д. Такая постановка вопроса имеет непосредственное отношение к нашей задаче, также связанной с переходом от языковых структур к метаязыковым (правда, иного свойства).

Важно подчеркнуть, что, несмотря на апелляции к психологическим факторам, единственной данностью для нас является текст. Опираясь на результаты конструктивно-синтаксического анализа, мы стремимся выделить из текста последовательность пропозиций и отобразить их на метаязыковую структуру. К сожалению, работа осложняется тем, что, помимо нехватки описаний, ориентированных на компьютерную обработку языка (что характерно и для предыдущих этапов), практически отсутствуют соответствующие практические разработки.

Если все же вернуться к литературе по рассматриваемым вопросам, следует отметить, что на данном этапе работы важными оказываются оба аспекта семантических исследований — как традиционный семасиологический (от формы к содержанию), так и ономасиологический (от содержания к форме), причем послед-

ний представлен в лингвистических исследованиях значительно более скупой, чем первый. В частности, весьма полезны описания грамматики русского языка, организованные вокруг тех или иных семантических полей: упомянем в этой связи серию коллективных монографий «Теория функциональной грамматики» под редакцией А. В. Бондарко, а также детальное описание лексико-синтаксической категории места в [Всеволодова, Владимирский 2009].

ОБЩАЯ КОНЦЕПЦИЯ ИНФОРМАЦИОННОЙ СТРУКТУРЫ ТЕКСТА

Построение информационной структуры текста начинается с создания двух объектов, релевантных для нее в целом и соответствующих, во-первых, фигуре повествователя и, во-вторых, самому тексту. Последнее необходимо для обработки разнообразных метакommunikативных элементов в тексте (ср. [Макаров 2003: 197–201]). Что касается фигуры повествователя, она требуется для того, чтобы каждое предложение (точнее, пропозиция) было привязано к источнику как сказанное либо повествователем, либо персонажем (третий вариант: цитата, атрибутируемая, как правило, некоторому внешнему для повествования лицу).

Разработанная нами концепция информационной структуры текста предполагает выделение двух основных типов данных — объектов и признаков. В целом это разграничение соответствует самому общему (хотя и известному своей неоднозначностью) делению всего того, что поставляет пищу нашему сознанию, на предметы и признаки.

Класс объектов включает живых существ (*Петя, писатель, дельфин*), неодушевленные физические предметы (*дом, трамвай, игрушка*), вещества (*молоко, песок*), ментально-информационные объекты (*адрес, доклад*), пространственные объекты, включая географические (*озеро, гора, дорога, Москва, Эверест*) и временные объекты (*вчера, 10 минут назад, I тыс. до н.э., 5 ноября 2017 г.*). Пространственные объекты должны дополнительно учитываться при построении пространственной схемы текста, отражающей взаимное расположение упомянутых в тексте объектов, их перемещение, ориентацию и пр.; причем для географических объектов указываются их координаты. Что касается временных объектов, они служат опорными точками при выявлении темпоральной структуры текста. Это со-

вершенно особенный вид объектов, так что многое, о чем говорится ниже, к таким объектам неприменимо.

Признаки могут характеризовать отдельные объекты или связывать их между собой, а также характеризовать другие признаки (*высокий дом, Петя спит, Петя купил книгу у Васи за 100 рублей, сугубо положительный, быстро бежать*). В отличие от объектов, имеющих пространственную привязку, признакам присущи темпоральные характеристики, которые должны быть отражены в специальном формате (подробнее см. здесь ниже).

Признаки бывают одноместные и многоместные. Признак считается одноместным, если он приписан одному объекту информационной структуры (даже если тот сам по себе является множественным), ср. *умный ребенок; умные дети; рогатый скот; Петя спит*. Мноместные признаки связывают между собой два и более объекта (*Петя учит детей математике; необходимые для работы навыки; разница во времени между Москвой и Камчаткой*). И те, и другие могут обозначать как постоянные признаки (*древний город стоит на берегу реки, грузовой лифт вмещает десять человек*), так и переменные (*голодный ребенок спит на полу, молодая листва шелестит на ветру*).

Признак в совокупности с соответствующими объектами образует пропозицию. Таким образом, согласно нашей интерпретации признака, мы видим пропозицию не только в классических случаях типа *Петя читает книгу, Петя спит*, но и в словосочетаниях *умный Петя, интересная книга* (ср. [Шмелев 1976: 71; Гак 1986: 54])³.

Принципиально важным мы считаем раскрытие так называемых свернутых пропозиций. Прежде всего, речь идет о номинализациях — отглагольных существительных со значением действия (*улучшение, рост, помощь, покраска*). В отличие от личных глаголов, номинализации нередко характеризуются неопределенностью в отношении субъекта действия, а также его временных характеристик, что приводит к незаполненным позициям актантов и дополнительным сложностям при проекции на ось времени⁴. Пустые позиции

³ Строго говоря, можно пойти и дальше, рассматривая в качестве пропозиций соответствующие экзистенциальные пресуппозиции: *Петя существует, книга существует*, однако вряд ли это имеет практический смысл. Появление того или иного объекта в тексте непременно влечет за собой его включение в информационную структуру независимо от его экзистенциального статуса в реальном или воображаемом мире.

⁴ Ср. [Серио 1999].

актантов должны восполняться на основе анализа контекста, но говорить о единой процедуре, похоже, здесь не приходится. Что касается временной отнесенности действия, выраженного номинализацией, и его статуса (реальный / возможный / несуществующий), эти параметры обусловлены грамматическими характеристиками и лексическим значением включающего предиката. Ср.: *в прошлом году обозначился (планировался, провалился) рост производства / в текущем году мы наблюдаем (ожидаем) рост производства / в следующем году будет происходить (намечается, предполагается) рост производства*. Более подробно проблема раскрытия свернутых пропозиций рассматривается в специальном разделе ниже.

Признаковые существительные типа *ум, храбрость, белизна, округлость* также образуют пропозицию (ср. *храбрость солдат, белизна снега*). Их употребление без зависимого существительного, по сути означает, что объект имеет универсальную референцию, ср. *Ум хорошо, а два лучше; Красота спасет мир*.

В информационной структуре признаки связаны между собой различными отношениями — временными, причинно-следственными, условными и т. д. Объекты также могут быть связаны между собой непосредственно, а не только через признаки в тексте (например, *комната — окно, сад — яблоня, поезд — пассажиры* и т. д.).

Такова общая концепция информационного представления. Обратимся теперь к процедуре преобразования синтаксических структур в метаязыковые.

ОТ СИНТАКСИЧЕСКИХ СТРУКТУР К ИНФОРМАЦИОННОМУ ПРЕДСТАВЛЕНИЮ ТЕКСТА

На выходе конструктивно-синтаксического анализа текст представлен в виде цепочки ПСДО предложений — от первого предложения текста вплоть до последнего. Процесс преобразования также идет последовательно слева направо: формирование информационной структуры текста начинается с ПСДО первого предложения, а ПСДО последующих предложений ее дополняют и развивают. Рассмотрим, что собой представляют элементы отдельно взятого ПСДО с точки зрения введенного выше разграничения на объекты и признаки.

Группа подлежащего (П) может содержать как предметную, так и признаковую лексику, соответственно, в информационной струк-

туре требуется создавать как объекты (ср. *мальчик, дерево, стол, линия*), так и признаки (ср. *ум, наводнение, говорить*). В самом простом случае, когда группа подлежащего образована единичным существительным типа 'живое существо' или 'предмет', дело ограничивается созданием в информационной структуре соответствующего объекта⁵. Исключением являются имена реляционной семантики (см. ниже).

Несколько сложнее обстоит дело, если группа подлежащего состоит из двух и более предметных существительных. Тогда наряду с введением новых объектов должны быть эксплицированы связи между ними, ср. *дом Маши, ручка двери, подарок для Пети, поезд из Москвы в Варшаву* и т. д. Вообще, все антропонимы в информационной структуре представлены объектом «человек» с приписанными ему признаками, в числе которых могут быть имя, профессия (*сторож, инженер*), возраст (*старик, подросток*), пол (*девушка, мужчина*), физические, интеллектуальные и прочие характеристики (*слепой, блондинка, дурак, гений, хам, истеричка*), социальные роли (*читатель, пассажир*) и пр.⁶ Так, если группа подлежащего представлена словосочетанием *фильм Тарковского* в информационной структуре создаются два объекта — «фильм» и «человек», при этом последнему приписывается признак «имя: Тарковский». Кроме того, эти два объекта необходимо связать между собой определенным отношением (проблема интерпретации смысловых отношений между членами генитивной конструкции рассматривалась в предыдущей главе). Реляционные имена (*мать, сосед, начальник, одноклассник*), совмещающие функцию идентификации объекта с информацией о его отношении к другому объекту (ранее упомянутому или имплицитруемому), предполагают одновременное добавление и новых объектов, и новых связей. Когда второй объект эксплицирован (*сосед Пети, друг Надиного мужа*)⁷, ситуация более простая. Если же группа подлежащего представлена одиночным реляционным существительным, в качестве второго объекта (своеобразной точки отсчета) выбирается текущий говорящий — повествователь или субъект прямой речи (персонаж). Далее, если мы имеем дело с субъек-

⁵ Кстати говоря, именно объекты образуют костяк информационной структуры — в отличие от синтаксической структуры, где ядром является предикат.

⁶ Из примеров видно, что многие антропонимы требуют квалификации одновременно по нескольким признакам.

⁷ Очевидно, что второй объект, в свою очередь, может заключать отсылку к третьему объекту и т. д.

ектом прямой речи, он, как правило, уже введен в информационную структуру, и референция слов реляционной семантики по умолчанию определяется относительно него. В то же время в ряде случаев (например, в вопросительных предложениях) в качестве второго объекта следует рассматривать фигуру адресата, ср. такие типовые реплики, обращенные взрослым к ребенку, как *Мать-то небось волнуется?* Если же реляционное существительное употреблено вне прямой речи, а повествование ведется от первого лица, отсылка осуществляется к повествователю как объекту с особым статусом (см. выше). Для повествования от третьего лица употребление одиночных реляционных существительных вне прямой речи нехарактерно.

Группа подлежащего может включать существительные, которые в зависимости от контекста выражают предметное или признаковое значение (*врач, преступник, оппозиционер* и т. п.). Соответственно, разработаны правила их интерпретации в различных контекстах, ср. *Врач пришел только вечером* vs. *Петя — врач/учится на врача/работает врачом*.

Если группа подлежащего содержит имена сугубо признаковой семантики (*последний фильм Тарковского, просмотр фильма Тарковского*), соответствующим объектам приписываются признаки, что в совокупности образует пропозицию, пусть даже и с частично незаполненными позициями. Так, в предложении *Просмотр фильма Тарковского занял три часа* номинализация *просмотр* разворачивается в пропозицию «просмотреть (кто? что? — фильм Тарковского)» с пустой позицией субъекта действия, которая предположительно будет заполнена на основании анализа соседних предложений.

Если пропозиция образована исключительно лексемами с признаковой семантикой (*Атака встретила ожесточенное сопротивление; Курить — здоровью вредить; Утром возникает желание позавтракать*), незаполненные позиции неизбежны, однако их интерпретация может быть различной. В первом примере выраженная подлежащим номинализация отсылает к конкретному действию, субъект которого должен быть определен по контексту. Во втором примере подразумевается универсальный субъект, что вообще характерно для одиночного инфинитива в роли подлежащего (так же, как и для одиночного существительного признаковой семантики, ср. *Бедность — не порок, Чистота — залог здоровья*). Что касается третьего примера, то он может иметь оба упомянутых прочтения. Таким образом, можно сформулировать четкие правила интерпретации применительно к одиночным инфинитиву и существительному

му признаковой семантики и тем самым сузить круг неоднозначных случаев.

Перейдем теперь к сказуемому (С). Сказуемое всегда представляет собой признак (или, по крайней мере, включает его, как в предложениях типа *Это была лисица*), что приводит к численному увеличению связей. Связи соединяют один или более объектов с признаком (*Петя болеет*, *Маша — филолог*, *Петя пригласил Машу в театр*). В последнем случае связи маркируются, для того чтобы различать реализуемые семантические валентности (субъект, объект, инструмент и т. д.). Если в качестве актантов выступают признаковые существительные (**Пожар** на атомной электростанции привел к **гибели** людей), сказуемое устанавливает соответствующие связи между признаками (*пожар*, *привести*, *гибель*), а те, в свою очередь, связываются с объектами (*пожар — атомная электростанция*, *гибель — людей*).

Если предложение содержит однородные подлежащие, информационная структура претерпевает различные изменения в зависимости от типа сказуемого. Возможны следующие ситуации:

- 1) одинаковые признаки приписываются каждому из объектов / признаков (*Рододендроны и бугенвиллеи не растут на севере*; *Пушкин и Лермонтов — гениальные поэты*; *Наводнение и пожар — это страшно*);
- 2) объекты / признаки соединяются посредством одного признака, выражающего взаимное действие или взаимное отношение (*Петя и Маша поцеловались* / *Петя с Машей поцеловались*; *Гений и злодейство несовместимы*);

Заметим, что примеры типа *Петя и Маша пошли в кино*, допускающие двоякую трактовку (совместное действие или самостоятельные поступки двух людей по отдельности), предполагают построение двух альтернативных структур для учета обоих вариантов.

Дополнения (Д) — это актанты глагола. Как и подлежащее, дополнение может иметь разветвленную структуру и включать слова предметной и признаковой семантики, причем первые не могут подчиняться себе вторых (*резать хлеб*, *сложить вещи в сумку отца*)⁸. Признаковые слова могут быть либо термами (*бороться с наводнением*), либо подчиняться себе другие признаковые и предметные имена (*противиться желанию спать*, *велеть принести чай*, *осужден за покушение на получение взятки*). Помимо этого, дополнение

⁸ Это возможно разве что в фигуральном смысле, ср. *хлеб изгнания*.

может представлять собой сентенциальный актанта (*Начальник распорядился, чтобы секретарша принесла кофе*). Все эти случаи получают разную интерпретацию в терминах объектов и признаков информационной структуры.

Так, предметные существительные в группе дополнения приводят к созданию новых объектов; если их несколько, то синтаксический хозяин (*сумка отца*) связывается с признаком соответствующим отношением (*сложить в сумку*). Наличие в группе дополнения определений (*сложить в новую сумку, прочитать интересную книгу*) означает прикрепление к создаваемым объектам соответствующих признаков и не меняет коренным образом сути построения.

Другое дело, если сказуемое представляет собой «включающий предикат» [Богданов 1977] и в группе дополнения есть слова предикатной семантики. В ряде публикаций [Семантика и синтаксис... 1981; Типология конструкций... 1985] было показано, что предикаты второго и более высоких порядков отнюдь не редкость в русском языке, к ним относятся фазовые, модальные и многие каузативные глаголы, ср. *начать репетировать/репетицию, должен уехать, заставил извиниться, пригласил совместно отпраздновать наши дни рождения*. Есть и такие глаголы, которые способны замещать свои валентности существительными как предметной, так и предикатной семантики — таковы, к примеру, глаголы зрительного восприятия и оценки [Арутюнова 1976: 136–138]. Как бы то ни было, мы всегда руководствуемся тем правилом, что в информационной структуре предикат должен быть развернут в пропозицию, даже если часть его мест не заполнена. В рассматриваемой ситуации дело осложняется необходимостью определения кореферентных связей между актантами включающего и включенного предикатов. Эта нетривиальная задача была подробно рассмотрена в статьях [Козинский 1985; Шаляпина 1985]. Заметим, что предложения с сентенциальными актантами, как правило, не создают проблем такого рода.

Обстоятельственные детерминанты (О) в зависимости от своего типа требуют разных построений в информационной структуре. Так, обстоятельства места, как правило, требуют создания новых объектов (*на борту самолета, около дома, в мире*), но могут быть представлены событийными именами со скрытой предикативностью, которые должны быть развернуты в пропозицию (*на конференции*). Обстоятельства времени (*во вторник, до Нового года, в период между двумя мировыми войнами*) предполагают создание временных объектов в виде точек, интервалов и т. д. на оси времени.

Конкретная временная локализованность, однако, может замещаться пропозицией: так, выражения *к празднику (город украсили флагами), до съезда (осталось пять дней)* содержат скрытую пропозицию, которая (как всякая пропозиция) имеет определенные координаты во времени и пространстве, но эксплицитных указаний на этот счет текст может и не содержать.

Что касается других видов обстоятельственных детерминантов (цели, причины и т.д.), они по своей сути всегда являются пропозициями, даже если представлены предложно-падежной группой, ср. *из-за дождя* — ‘из-за того, что идет/шел дождь’, *для Петиного выздоровления* — ‘для того, чтобы Петя выздоровел’, *несмотря на отсутствие денег* — ‘несмотря на то, что (у некоторого субъекта) отсутствуют/отсутствовали деньги’. Поскольку такие выражения являются скрытыми пропозициями, возникает уже упоминавшаяся проблема их недостаточной информационной эксплицитности. Незаполненные места предиката могут порождать неоднозначность прочтения. Кроме того, в подобных конструкциях в принципе должны употребляться предикатные существительные (как в примерах выше), но на их место нередко встают существительные с предметным значением, ср. *из-за машины, ради собаки, благодаря Николаю Ивановичу* и т.п. Специфика контекста диктует событийное прочтение, иными словами, подобные существительные также должны разворачиваться в пропозицию, однако проблема информационной недостаточности здесь еще более серьезная, чем в случае номинализаций.

НАКОПЛЕНИЕ ИНФОРМАЦИИ: КОНТРОЛЬ ЗА ОБЪЕКТАМИ И ПРИЗНАКАМИ

Выше мы рассмотрели возможные соотношения между синтаксической структурой отдельного предложения, выраженной набором (-ами) ПСДО, и элементами информационной структуры. Излишне говорить, что обработка каждого следующего предложения (точнее, ПСДО) не только усложняет уже описанные процессы преобразования, добавляя все новые объекты и признаки, но и ставит задачи принципиально иного уровня, связанные с автоматическим анализом текста как единого целого. Нам неизвестны какие-либо лингвистические модели, в которых эта идея была бы осуществлена

в полном объеме⁹. Наши разработки в данном направлении также пока носят фрагментарный характер.

В самом широком плане речь идет о выявлении содержательных связей (локальных и глобальных) между предложениями, которые в совокупности позволяют определять смысловую направленность текста, его тему. Эта задача внимательно исследовалась в рамках лингвистики текста. Как известно, надежды ученых на то, что удастся выделить конечный набор типов связей и способов их выражения, не оправдались. Более того, полемически заостряя ситуацию, Т.М.Николаева [1978] продемонстрировала, что формальная связанность текста (когезия) в общем случае не является ни необходимым, ни достаточным условием его содержательной связанности (когеренции).

Однако когда мы имеем дело с осмысленным текстом, категоричность этого утверждения заметно снижается: наличие когезивных маркеров все же несет определенную информацию о когеренции текста, но ее недостаточно для того, чтобы определить все связи между соответствующими объектами и/или признаками. Следует также подчеркнуть, что выявление средств когезии в том или ином тексте и извлечение из них информации становятся самостоятельной задачей; из разнообразных языковых механизмов когезии (см., например, [Halliday, Hasan 1976]), пожалуй, только лексический повтор и коннекторы не представляют большой сложности с точки зрения автоматической обработки. Что же касается вычисления анафорических и катафорических отсылок, восстановления эллипсиса, установления кореферентных связей между именными группами, определения тема-рематических прогрессий, все это отдельные вопросы, требующие фундаментальной проработки.

Совершенно иной подход к проблеме выявления связности текста был предложен специалистом в области искусственного интеллекта Р.Шенком и его коллегами. Стремясь построить психологически достоверную лингвистическую модель, работа которой будет «насколько возможно, близко соответствовать процессам, происходящим в человеческом мозгу» [Шенк 1980: 20], авторы исходили из известного в когнитивной психологии феномена ожидания. Суть

⁹ Так, знаменитая модель «Смысл ↔ Текст» на деле работает не с целостным текстом, а с отдельными предложениями [Апресян, Цинман 1998]. Модель Р.Шенка, о которой будет сказано ниже, не была доведена до уровня практической реализации.

его заключается в следующем: по ходу восприятия текста (устного или письменного) человек непроизвольно строит ожидания (предсказания), касающиеся вероятного продолжения текста. Предсказания могут оказаться неверными, но в большинстве случаев они оправдываются. В потоке текста эти предсказания эксплицируют неявные смысловые связи между соседними предложениями, относящиеся к области «наивной психологии» — того, как человек понимает окружающий мир, что побуждает его действовать и т. д.

Моделируя механизм предсказаний, авторы вводят понятие концептуального умозаключения, призванное восполнять подобные «пропущенные звенья». Они выделяют характерные типы умозаключений, которые отражают наиболее регулярные частотные связи между соседними предложениями в тексте и при отсутствии эксплицитных маркеров связи способны помочь компьютеру понимать текст как связанное целое. Подход Шенка служит удачным примером интеграции достижений когнитивной психологии и теории искусственного интеллекта; к сожалению, его реализация оказалась фрагментарной, и модель не получила дальнейшего развития. В то же время идея моделирования концептуальных умозаключений представляется весьма плодотворной в контексте экспликации смысловых связей в тексте. Она могла бы найти свое воплощение в энциклопедическом компоненте, разработка которого является на сегодняшний день насущной задачей.

Обратимся теперь к нашей конкретной задаче — созданию информационной структуры по тексту — и рассмотрим проблему смысловой связанности в этом аспекте. Поскольку элементами информационной структуры являются объекты и признаки, речь идет об определенном отслеживании, контроле за происходящим с теми и другими по мере того, как обрабатываются все новые ПСДО следующих друг за другом предложений. Применительно к объектам могут осуществляться следующие операции:

- создание объекта (операция применяется не только при каузации возникновения нового объекта типа *построить дом, сварить суп, выкопать яму*, но и вообще всякий раз, когда в тексте упоминается объект, на тот момент отсутствующий в информационной структуре¹⁰);

¹⁰ Иными словами, данная операция применяется всегда, когда упоминаемый объект не удается наверняка связать кореферентной связью с уже имеющимися

- создание объекта и его связывание с уже созданными объектами (ср. *Маша — ее муж — его машина — автомастерская, в которой ее чинят — ...*);
- маркировка объекта как несуществующего (начиная с определенного момента), ср.: *человек умер, дом снесли, пирог съели*.

Контроль за объектами в информационной структуре осуществляется путем приписывания им референциальных индексов. Каждому новому объекту либо присваивается новый индекс, либо производится выбор из ранее приписанных индексов. Последнее означает, что объект уже существует в информационной структуре под другим именем, и в этом случае ему просто приписывается еще одно, дополнительное, имя. Эта процедура технически проста, но в теоретическом плане вызывает много вопросов, поскольку затрагивает целый спектр проблем, касающихся референциальных отношений в связном тексте, в том числе: вычисление анафорических отсылок, восстановление эллипсиса, определение референциального статуса именных групп, идентификацию кореферентных лексических единиц. Каждая из них является весьма сложной и многоплановой в аспекте автоматического моделирования обработки текста.

Так, чрезвычайно широкая литература (в особенности зарубежная) посвящена вопросу отыскания антецедентов местоимений (и других близких к ним слов-заместителей) в связном тексте; решение этой задачи, как правило, основывается на использовании вероятностно-статистических методик. Проблема определения референциального статуса имен неоднократно затрагивалась в трудах логиков (преимущественно на материале английского языка); применительно к русскому языку она глубоко и всесторонне анализируется Е. В. Падучевой [1985]. Безартиклевые языки в этом плане представляют еще больший простор для неоднозначности, чем языки с артиклями¹¹.

Впрочем, и в тех и в других языках референциальный статус отдельной именной группы в составе предложения может быть неясен даже читателю, не говоря уже о системе искусственного интеллекта,

объектами. В результате количество объектов в информационной структуре может заметно превышать число соответствующих референтов.

¹¹ Установлению референции может помочь учет некоторых дискурсивных маркеров, однако работы в этом направлении только начинаются (ср. [Борисова 2014]).

так что его определение требует привлечения более широкого контекста, ср. *Скорпион похож на кузнечика* (скорпион вообще или данный скорпион? на кузнечика вообще или на данного кузнечика?), *У кошки глаза зеленые* (у данной кошки или у всех кошек?), *Иван подрался с милиционером* (с конкретным милиционером или каким-то представителем данной профессии?)¹². Если поясняющий контекст отсутствует, в информационной структуре просто создаются соответствующие объекты: *скорпион, кузнечик, кошка, Иван, милиционер*.

В то время как определение анафорических отсылок (и отчасти восстановление опущенных элементов в эллиптических предложениях) базируется на анализе относительно близкого контекста, установление кореференции имен требует семантического анализа в более крупном масштабе. Без решения этой задачи невозможно выявить содержание текста. Использование различных признаков для номинации, в частности, широко распространено среди антропонимов. Так, один и тот же человек может быть обозначен как *лесник, старик, хитрец, толстяк, сосед...* Однако этот ряд контекстуально взаимозаменяемых существительных имеет свои ограничения, и продолжить его словами *подросток* и *тракторист* невозможно (ср. [Шмелев 1973: 106])¹³.

Подобные примеры естественным образом наводят на мысль о необходимости описания совместимости/несовместимости имен с точки зрения кореференции. Именно эта идея была положена в основу гипотезы референциального отождествления имен (см. [Рубашкин 1989: 126–143]). Заметим, что даже если бы эта гипотеза подтвердилась, задача определения кореференции в тексте не была бы решена в полном объеме: ведь помимо отдельных слов, совместимость которых можно задать в словаре, в данное отношение способны вступать сколь угодно сложные словосочетания, не поддающиеся исчислению. Однако, как вынужден констатировать сам автор, его гипотеза была опровергнута контрпримерами. Более того, оказалось, что даже лексический повтор не гарантирует референциального тождества, ср.: *Из пункта А в пункт Б выехала машина, а через 5 часов выехала машина из пункта Б в пункт А* [Рубашкин 1989: 140]. От себя добавим, что не только машины, но и сами пункты,

¹² Примеры Е. В. Падучевой.

¹³ В более широкой перспективе данный вопрос выводит нас на проблему вариативной интерпретации действительности [Баранов, Паршин 1986], а также когнитивные исследования категоризации мира и схематизации описываемой ситуации [Lakoff, Johnson 1980: ch. 24; Talmy 1983; Lakoff 1987; Талми 1999].

условно названные А и Б, могут быть разными: не так уж и сложно найти на карте населенные пункты с одинаковым названием.

Что касается признаков, они всегда приписываются каким-то объектам как опорным точкам информационной структуры. Самая простая ситуация связана с добавлением одноместного признака к ранее созданному или впервые упомянутому объекту (*Маша ленива; Маша спит*). В случае множественных признаков объекты связываются друг с другом, что может вовлекать уже имеющиеся в информационной структуре объекты, либо требовать создания новых (*Маша любит Петю; Маша разговаривает по телефону*).

Важный вопрос состоит в том, что происходит с признаками, которые были ранее приписаны тому или иному объекту. Оказывается, они могут сохраняться (*маленькая* девочка смутилась и покраснела; *белый* хлеб зачерствел; *тетрадь в клетку* кончилась; *Вася сел* на скамейку и углубился в чтение газеты; *Иван Иванович шел* по улице и думал о друге), но могут также претерпевать изменения (*яркая* краска потускнела и осыпалась; *шквальный* ветер понемногу ослабевал) вплоть до отмены (*бедняк* разбогател; *свежий* хлеб зачерствел; *Петя долго листал* газету, потом выключил свет и уснул). Пока не вполне ясна природа взаимодействия между старыми и новыми признаками. Хотелось бы каким-то образом систематизировать наши представления о том, в каких случаях признаки изменяют приписанную ранее информацию, а в каких просто ее добавляют. Очевидно, что на это влияет характер обоих признаков — и старого, и нового. Этот вопрос требует углубленного изучения, но уже сейчас можно предположить, что в конечном итоге дело упирается в наши знания не только о языке, но и о мире.

Вопрос о контроле за признаками выводит нас на задачу определения временной структурой описываемых в тексте событий, которая рассматривается в следующем разделе.

ПРЕДСТАВЛЕНИЕ ИНФОРМАЦИИ О ПРОСТРАНСТВЕ И ВРЕМЕНИ

Контроль за объектами, в частности, предполагает отслеживание их местонахождения и перемещения в пространстве. Эта задача включает целый комплекс проблем¹⁴, с которыми в полном объеме

¹⁴ Достаточно полное представление о них можно составить по книге [Кандрашина, Литвинцева, Поспелов 1989].

не смог справиться на настоящий момент ни один коллектив разработчиков моделей искусственного интеллекта. Основная причина, по-видимому, заключается в отсутствии единой системы, в рамках которой может осуществляться описание местонахождения объекта средствами естественного языка. Ведь указание на пространственное расположение может определяться по абсолютной (*в Санкт-Петербурге*) или относительной шкале (*неподалеку от дома*), причем в последнем случае отсчет может осуществляться как в экзосцентрической, так и в эгоцентрической системе координат (ср. *В ста метрах на запад от железнодорожной станции Репино* vs. *В ста метрах к западу от моего дома*). Из этого следует, что даже если можно было бы загрузить в модель подробные географические сведения в том или ином формате, это не решило бы проблему интерпретации локативных выражений.

Если попытаться рассмотреть разные типы текстов в указанном ракурсе, мы обнаружим, например, что в коротких новостных текстах преобладает абсолютное шкалирование, хотя нередким является и относительное (экзосцентрическое). Для художественных произведений, напротив, характерно использование относительных шкал; лишь начало описываемых событий может быть привязано к известному, закреплённому в сетке координат, месту. Литературное повествование разворачивается преимущественно в эгоцентрической системе координат, организованной вокруг рассказчика или персонажа.

В связи с этим актуальным становится вопрос об интерпретации дейктических выражений и учете такого параметра, который в различных работах получил название перспективы, или точки зрения, или наблюдателя. Несмотря на традиционное внимание к теме дейксиса со стороны как отечественных (Ю. Д. Апресян, Е. В. Падучева и др.), так и зарубежных (Ч. Филмор, Дж. Лайонз и др.) ученых, заметно усилившееся в последние десятилетия в связи со становлением когнитивной лингвистики и интересом к языковой концептуализации мира [Talmy 1983; Langacker 1991; Яковлева 1994; Булыгина, Шмелев 1997; Логический анализ языка... 1997; Levinson 2003], исследования в этой области, как правило, не носят прикладного характера. Компьютерная лингвистика, по крайней мере отечественная, мало продвинулась в данном направлении.

Правильная интерпретация дейктических выражений важна также для успешной автоматической обработки прескриптивных текстов (инструкций, правил, предписаний, руководств, кулинар-

ных рецептов и др.), многие из которых вообще лишены привязки к абсолютной шкале и широко задействуют эгоцентрическую систему координат. Пространство, к которому осуществляется референция, часто является весьма небольшим (по сравнению, скажем, с художественными и публицистическими текстами), ограниченным рамками того или иного помещения (например, метрополитена), устройства или механизма (например, бытовой техники).

Ограниченность пространства характерна и для дескриптивных текстов, хотя масштаб здесь может быть различным (очевидно, что объектом описания может стать объект сколь угодно малый и, наоборот, сколь угодно большой). Но этот объект заранее задан (как правило, отражен в заголовке текста), и за его границы текст обычно не выходит. Тем не менее это может случиться, если автор сопоставляет описываемый объект с каким-то другим, в некотором роде аналогичном, но расположенном в другом месте. Например, в описание какого-нибудь примечательного здания или сооружения могут быть включены отсылки к другим постройкам того же архитектора, инженера и т.п. В дескриптивных текстах часто встречается первоначальная привязка к абсолютной шкале (ср., например, путеводители). В дальнейшем, правда, такие тексты опираются на относительные шкалы — как экзо-, так и эгоцентрические. Все сказанное, однако, не распространяется на дескриптивные фрагменты, включенные в литературный нарратив (описания персонажей, зарисовки природы и др.).

Заметим, что в русском языке, по-видимому, нет таких локативных выражений (от предлогов до сколь угодно сложных оборотов), которые были бы предназначены исключительно для референции к какой-то одной (из пяти возможных) системе отсчета: абсолютной (которая всегда экзоцентрична), относительной или размытой (последние две могут быть как экзо-, так и эгоцентрическими). В текстах на естественном языке данные системы отсчета причудливо переплетаются, серьезно затрудняя попытки пространственного моделирования. Добавим к этому, что заключенная в тексте информация может касаться различных локативных характеристик — положения одного объекта или взаимного расположения двух и более объектов, расстояния между ними, направления движения, ориентации. Таким образом, необходимо учитывать ряд взаимосвязанных параметров. А если речь идет о динамической ситуации, пространственные характеристики необходимо соотносить с временными координатами, автоматическое определение которых осложнено действием целого ряда факторов (см. ниже).

Дополнительные трудности при автоматической обработке выражений с пространственной семантикой возникают из-за широкого использования в естественном языке размытых систем отсчета [Кандрашина, Литвинцева, Пospelов 1989: 198], которые по своей природе могут быть как экзoцентрическими, так и эгоцентрическими, ср. *недалеко от Петербурга* vs. *недалеко от машины*. Неопределенность локализации, заключенная в выражениях типа *вблизи от*, *слева от*, *наискосок от* еще более усиливается за счет того, что оценка расстояния зависит от величины соответствующих предметов: одно дело *чашка рядом с тарелкой*, другое дело — *город рядом с границей*. Имеется и более тонкий момент, связанный с прагматическим параметром «фоновые допущения и ожидания» [Langacker 1988: 67]: скажем, близость работы к дому по-разному будет оцениваться жителями Москвы и маленького провинциального городка.

Известная размытость, неточность семантики присуща не только наречиям и предлогам с пространственной семантикой. Возьмем, к примеру, слово *окрестность*: какую территорию можно назвать *окрестностью* чего-либо? Очевидно, что однозначный ответ здесь невозможен. Во-первых, как и в примерах выше, важен размер так называемой точки отсчета (она же — фон, или ориентир): словосочетание *в окрестностях деревни* предполагает круг с радиусом едва ли больше десяти километров (а скорее даже меньше), в то время как выражение *в окрестностях мегаполиса* способно обозначать гораздо большую территорию. Во-вторых, понятно, что проведение границ «окрестности» при помощи циркуля не соответствует узусу: так, если город расположен близко к границе с другой территориальной единицей (штатом, республикой, страной), его *окрестности* едва ли включают часть этой единицы. Наконец, в данном понятии присутствует смысловой компонент, связанный с доступностью (где-то не более двух-трех часов пути), поэтому можно предположить, что границы «окрестности» с развитием транспортных средств постепенно расширяются.

Напоследок упомянем неоднозначность интерпретации сложных локативных оборотов, включающих запятую, ср. *выступал в Москве, в Большом театре* vs. *гулял на Марсовом поле, на Каменном острове*). Соотношение обозначенных мест, разделенных запятой в приведенных примерах различно: в первом случае второе «вложено» в первое, во втором случае они не пересекаются. Подобная неоднозначность может возникать как в позиции актанта при глаголах с валентностью места, так и в обстоятельственных детерминантах,

относящихся к предложению в целом. Правильная интерпретация таких контекстов невозможна без введения в лингвистический процессор энциклопедического компонента, содержащего описание знаний о мире.

* * *

Если объекты характеризуются положением в пространстве, то признаки — временными параметрами. Определение временной структуры текста на естественном языке является чрезвычайно сложной задачей, что обусловлено рядом факторов. Прежде всего, как говорилось выше, приписывание объекту нового признака само по себе ничего не говорит о сохранении или отмене ранее приписанных признаков. Вероятно, можно провести аналогию с совместимыми и несовместимыми номинациями объектов, хотя здесь мы находимся на еще более зыбкой почве.

Но основная, принципиальная проблема, с которой мы здесь сталкиваемся, — это та, что является единой для отображения пространственных и временных отношений. Мы имеем в виду невозможность построения единой модели, что обусловлено сосуществованием в текстах на естественном языке двух систем отсчета (как пространства, так и времени) — абсолютной и относительной [Кандрашина, Литвинцева, Поспелов 1989: 66–75; Попов 2004: 86, 137].

Применительно ко времени абсолютная система упорядочивает события вдоль оси «прошедшее — настоящее — будущее», где настоящее ассоциировано с моментом речи, прошлое привычно располагается слева от него, а будущее — справа. Следует заметить, однако, что в соответствии со спецификой языковой картины мира настоящее должно моделироваться не как точка на шкале, а как некий интервал с размытыми границами (ср.: *Сейчас я обедаю/занят/разбираю документы/еду на работу/живу в Москве/работаю учителем в школе* и т. п.). Описываемые события на абсолютной шкале могут иметь жесткую привязку к календарному времени (*Стрелки часов были переведены 25 марта в 2 часа ночи по московскому времени*) или требовать дейктического прочтения (*Завтра мы поедem за город*). Примечательно, что настоящее (момент речи) обычно в тексте как таковое не фиксируется, а расположение тех или иных событий слева или справа от него определяется по грамматической форме глагола и обстоятельствам времени, в том числе дейктическим выражениям.

Относительная система отсчета описывает время протекания одного действия (состояния и др.) относительно времени другого действия (состояния). Соответственно выделяются три значения: прямое, предшествующее, последующее. Ср.: *Он появился в тот момент, когда я в нетерпении уже набирал номер его телефона* [прямое]; *Перед тем как идти в гости, мы плотно пообедали* [предшествующее]; *Я заснул около полуночи и проспал 10 часов* [последующее]. Приведенные примеры демонстрируют чистые случаи включения одного события в другое или, напротив, их разделенности; между тем довольно часто встречаются и пересечения интервалов, ср.: *Не успел я толком войти и раздеться, как раздался телефонный звонок; Заря уже занималась, когда они выступили в путь.*

Вообще, временные границы описываемых в тексте событий (как и, собственно говоря, самих событий в действительности) нередко размыты, неопределенны. Чтобы в этом убедиться, достаточно, например, попытаться определить начало (или конец) зари, сумерек, поездки за город (с момента начала сборов? надевания верхней одежды? выхода из квартиры?), рабочего дня, если он официально не задан жестким интервалом (с момента входа в здание? в рабочее помещение? начала конкретной трудовой деятельности?) и др.

Отметим также, что временные объекты весьма разнородны в том, что касается их масштаба (ср. *в эпоху неолита, во II тысячелетии до н.э., 25.09.2017, в 23:14 пятого августа 2000 г.*) и природы (в языке, наряду с обозначениями одного конкретного интервала, имеются темпоральные выражения, референтом которых являются регулярно повторяющиеся промежутки, ср. *по утрам, каждой осенью*).

Еще одна сложность заключается в неоднозначности референциальной отсылки некоторых темпоральных выражений. Так, *в понедельник* может относиться как к прошлому, так и к будущему понедельнику на абсолютной шкале времени, причем данную неоднозначность невозможно надежно разрешить опираясь на форму глагола, ср. *В понедельник я иду на концерт* [будущее] vs. *В понедельник я иду на концерт и вижу...* [прошлое]. Дейктический оборот *в этом году* за исключением приблизительной фиксации на шкале времени сам по себе не несет информации о том, относится ли описываемое событие к прошлому, настоящему или будущему.

Все эти факторы — недискретность внешнего мира, определенная размытость и субъективность его отражения средствами естественного языка, а также неоднозначность референции языковых

выражений — ставят под вопрос пригодность формально-логического аппарата для моделирования содержания текста.

Абсолютная и относительная системы отсчета времени, так же как и соответствующие модели пространственной локализации, в текстах весьма часто совмещаются. В том виде, как они охарактеризованы выше, эти системы способны обеспечивать лишь весьма грубое, упрощенное представление о временной структуре текста на русском языке. В них невозможно учесть все разнообразные аспекты протекания различных действий (процессов, состояний и пр.), выражение которых обеспечивается сложным взаимодействием между лексическим значением глагола, его аспектуальными характеристиками, а также семантикой различных обстоятельственных конструкций. Речь идет о таких параметрах, как различные способы действия, предельность / неопределенность, заполненность / незаполненность временного отрезка действием, единичность / повторяемость действия.

К сказанному следует добавить, что текст может содержать более одной версии событий (действительных или воображаемых) — как в прошлом (значение ирреалиса), так и в настоящем и в будущем. Описывающие их предложения могут опираться на абсолютную или относительную систему отсчета времени, а также на их комбинацию.

Как абсолютная, так и относительная временные ориентации имеют в русском языке специализированные лексические и грамматические средства выражения, однако соответствующие списки (см. [Попов 2004: 190–191]), разумеется, не являются исчерпывающими. Попытки моделирования временной структуры текста на русском языке дополнительно осложняются отсутствием так называемого согласования времен, которое в ряде других языков (английском, французском и т. д.) позволяет более надежно соотносить действия, состояния, отношения и др., выраженные личными и неличными формами глагола. В русском языке время осуществления действия может определяться по отношению либо к моменту речи (*Он придет вовремя*) либо к какому-нибудь другому моменту, в частности ко времени другого действия (*Он думал, что придет вовремя*); проблема состоит в том, что в системе форм времени эти различия не отражены. Иными словами, морфологическое время не отражает непосредственно «семантическое время» [Попов 2004: 191].

ОПРЕДЕЛЕНИЕ ВРЕМЕННОЙ СТРУКТУРЫ ТЕКСТА

Кратко охарактеризовав специфику отражения временных отношений в русском языке и трудности, возникающие при попытке осуществлять автоматический контроль за темпоральной структурой текста, постараемся теперь указать на отдельные подсказки, или зацепки, которые способны направлять этот процесс. В идеале, конечно, хотелось бы маркировать каждый признак некоторым индексом, соотнесенным с временной шкалой (или шкалами), однако на практике дело ограничивается пока обработкой отдельных, более или менее однозначных случаев¹⁵.

Начнем с так называемых прескриптивных текстов, или текстов-инструкций [Хорохордина 2013]. Объединяя широкий спектр речевых жанров (закон, указ, распоряжение, руководство, предписание, приказ, рекомендация, инструкция, рецепт, совет, просьба, порядок действий и др.), этот тип текста обладает характерными особенностями. С точки зрения временной организации интерес представляют следующие.

Во-первых, тексты (или их отдельные компоненты), которые непосредственно описывают некоторую процедуру, демонстрируют иконическое соотношение между порядком следования глаголов и последовательностью описываемых действий. При этом глаголы могут употребляться в форме повелительного наклонения, инфинитива или настоящего времени. Приведем типичные примеры:

***Поместите** продукты в духовой шкаф, **выберите** нужный режим духового шкафа и **отрегулируйте** температуру приготовления продукта. Нажимая кнопку, **выберите** режим «Окончание приготовления». Соответствующий индикатор начнет мигать. Затем **действуйте** следующим образом...*¹⁶

*Рыбу **очистить** от чешуи, **отрезать** голову и, не разрезая брюшка, **вынуть** внутренности; **удалить** из головы жабры, **промыть** холодной водой и **нарезать** рыбу поперечными кусками*¹⁷.

¹⁵ Отметим, что в литературе подробно рассмотрены временные соотношения, выраженные последовательностями «глагол СВ + союз и + глагол СВ», «глагол НСВ + союз и + глагол НСВ», «глагол СВ + союз и + глагол НСВ» в пределах одного предложения (см. [Маслов 1984; Падучева 1996: 362–364; Падучева 2016]). Нам неизвестны какие бы то ни было попытки выйти за рамки одного предложения.

¹⁶ Газовый духовой шкаф ЕОГ 23400: Руководство по эксплуатации.

¹⁷ Книга о вкусной и здоровой пище. М., 1952. С. 136.

Смешанную с содой муку **рубят** на столе с маслом. Желтки **растирают** добела с сахаром, **смешивают** с рубленой мукой и **замешивают** тесто, которое **выносят** на 10–15 минут на холод. Охлажденное тесто **раскатывают** в круглую лепешку и **укладывают** в форму или сковороду, смазанную маслом. Поверхность лепешки **посыпают** рублеными орехами, смешанными с сахаром (полстакана), и **заливают** взбитыми в крепкую пену белками, смешанными с сахаром (полстакана). Пирог **выпекают** при температуре 230–240 °С¹⁸.

Догнать и перегнать Америку!

Приди и проголосуй!

Независимо от формы глагола, описываемые действия обращены в будущее — относительно и момента речи (порождения текста), и момента его восприятия адресатом. Настоящее здесь не играет существенной роли: люди могут руководствоваться идентичными инструкциями в разные дни, месяцы и годы. Границы будущего неопределенные, поскольку обусловлены внешними факторами (скажем, принятием новых постановлений, отменяющих старые).

Исключение составляют бытовые директивы, прочно укорененные в «здесь и сейчас» и содержащие конкретные указания относительно действий адресата (как правило, в ближайшем будущем). Можно сказать, что бытовые прескриптивные тексты (скорее речевые акты), такие как просьба, совет, предсказывают возможность исполнения соответствующего действия лицом, к которому обращена директива. (В зависимости от формы глагола этот агент может мыслиться как единичный или множественный, включать или исключать субъекта речи.) В этом случае имеет смысл создавать в информационной структуре соответствующие объекты и признаки, правда, со специальным статусом (не фактивным, а вероятным). Это может оказаться полезным для восполнения недостающей информации (эллипсиса) и экспликации содержательной связанности текста.

Во-вторых, многие прескриптивные тексты включают перечень разрешений и ограничений, выражаемых при помощи отрицательных форм глагола, а также разнообразных модальных слов и выражений, таких как *(не) должен, (не) следует, можно, необходимо, (не)*

¹⁸ Даниленко М. П. Как приготовить дома кондитерские и другие изделия из муки, сладкие блюда, варенье, соки и припасы на зиму (пособие для молодых хозяек). М., 1961. С. 104.

допускается, (не) разрешается, (не) имеет право (-а) не наблюдается хронологического упорядочения соответствующих предикатов: все они одновременно действуют применительно к описываемой ситуации начиная с настоящего (момента порождения текста) вплоть до некой неизвестной точки в будущем. Например:

Гибкий шланг можно использовать при соблюдении следующих условий:

- шланг **не должен нагреваться** выше комнатной температуры (выше 30 °С длина шланга **не превышает** 1500 мм;
- в шланге **отсутствуют** участки сужения просвета;
- шланг **не натянут и не перекручен**;
- шланг **не соприкасается** с острыми кромками или углами;
- шланг **можно легко осмотреть** для проверки его состояния¹⁹.

В-третьих, следует отметить практически полное отсутствие глагольных форм прошедшего времени в прескриптивных текстах.

Наконец, нельзя не упомянуть об особенном употреблении форм настоящего времени при описании потенциальной типовой процедуры, локализованной в будущем (по отношению к моменту составления текста) — так называемом «настоящем предписания» [Кожин, Крылова, Одинцов 1982: 106]. В дополнение к приведенной выше выдержке из кулинарного рецепта сошлемся на тексты официально-делового стиля, где это встречается довольно часто. Вот характерный пример: *О причинении гражданину телесных повреждений в результате применения сотрудником полиции физической силы, специальных средств или огнестрельного оружия полиция в возможно короткий срок, но не более 24 часов, уведомляет близких родственников или близких лиц гражданина*²⁰.

Если не учитывать специфику языка официальных документов²¹, данный фрагмент текста будет (неправильно) интерпретирован следующим образом: объект 'полиция' в информационной базе соединяется с предикатом 'уведомлять', причем последний имеет статус 'реальный, текущий', который соответствует центральному значению форм настоящего времени и приписывается им по умолчанию. Заметим, что в других контекстах такая интерпретация будет верна,

¹⁹ Газовая варочная панель AEG: Инструкция по эксплуатации.

²⁰ Федеральный закон от 07.02.2011 № 3-ФЗ «О полиции», статья 19, п. 5.

²¹ Заметим, что в бытовом разговоре, художественном или публицистическом тексте та же мысль, скорее всего, была бы оформлена в виде условного предложения (*Если гражданин получил телесные повреждения... полиция обязана уведомить...*).

ср.: В данный момент полиция звонит родственникам и уведомляет их о... Это наглядный пример того, что выбор правильной интерпретации морфологических форм глагола зависит от стиля и жанра текста. Проблема в том, что описываемый лингвистический процессор не знает, какой текст он обрабатывает. Так возникает непростая задача автоматического определения типа текста, которой посвящен следующий раздел.

* * *

Возвращаясь к феномену иконической упорядоченности, нельзя не вспомнить о знаменитой статье [Labov, Waletzky 1967], авторы которой обратили внимание на то, что последовательность предложений (точнее, личных форм глаголов) в основной части (так называемом осложнении) коротких устных рассказов отражает порядок соответствующих действий. Применительно к русскому языку их наблюдения могут быть использованы при анализе последовательностей простых предложений²² в нарративном тексте, где сказуемые выражены личными глаголами совершенного вида. Рассмотрим в качестве примера следующий рассказ Д. Хармса:

*Однажды Антон Бобров **сел** в автомобиль и **поехал** в город.*

*Автомобиль **наскочил** на ломаные грабли.*

***Лопнула** шина.*

*Антон Бобров **сел** на кочку возле дороги и **задумался**.*

*Вдруг что-то сильно **ударило** Антона Боброва по голове.*

*Антон Бобров **упал** и **потерял** сознание.*

На основе анализа подобных примеров было сформулировано следующее рабочее правило: если в тексте употребляются исключительно глаголы совершенного вида в положительной форме и одинаковом грамматическом времени, характеризующие действия одного и того же персонажа, а эксплицитные маркеры темпоральных отношений отсутствуют²³, «по умолчанию» вступает в силу прин-

²² Появление сложноподчиненных предложений нередко нарушает иконический принцип отражения событий в тексте (ср. *Я срочно **выехал** в Москву, как только **получил** письмо*). Что касается сложносочиненных предложений, принцип обычно сохраняет свое действие, если части предложения (клаузы) соединены союзом *и*.

²³ В подобных случаях, как замечает А. В. Бондарко, «глагольный вид выполняет не только свойственные ему аспектуальные функции, но и функции темпоральные» [Бондарко 1971б: 45].

цип иконической интерпретации (ср. известный пример *Veni, vidi, vici* [Якобсон 1983: 107]).

Заметим, однако, что такие тексты встречаются весьма редко — не случайно чистый пример оказалось легче найти в литературе абсурда. В обычном художественном произведении иконичность сохраняется разве что на небольших отрезках текста, ср.:

*Утро застало Иванова осунувшимся, похудевшим... Он тихонько **пробрался** в гостиную, **схватил** газету и, **убежав** в спальню, **развернул** свежий газетный лист.*

*Через пять минут он **вбежал** в комнату жены и дрожащими губами **прошептал**...* (А. Аверченко. История болезни Иванова).

Примечательно, что даже если фрагмент повествования насыщен действием, в рассказ часто бывают вкраплены глаголы несовершенного действия, как, например, в отрывке ниже:

*Я **вздрыгнул**... Мне **почудился** скрип отворявшейся двери, потом легкий треск переломанного сучка. Я в два прыжка **спустился** с развалины — и **замер** на месте. Быстрые, легкие, но осторожные шаги явственно **раздавались** в саду. Они **приближались** ко мне. «Вот он... Вот он, наконец!» — **промчалось** у меня по сердцу. Я судорожно **выдернул** нож из кармана, судорожно **раскрыл** его — какие-то красные искры **закрутились** у меня в глазах, от страха и злости на голове **зашевелились** волосы... Шаги **направлялись** прямо на меня — я **стибался**, я **тянулся** им навстречу... **Показался** человек... боже мой! это **был** мой отец!*

*Я **тотчас** **узнал** его, хотя он весь **закутался** в темный плащ и шляпу **надвинул** на лицо. На цыпочках **прошел** он мимо. Он не **заметил** меня, хотя меня ничто не **скрывало**, но я так **скорчился** и **сжежился**, что, кажется, **сравнился** с самой землей. Ревнивый, готовый на убийство Отелло внезапно **превратился** в школьника... Я до того **испугался** неожиданного появления отца, что даже на первых порах не **заметил**, откуда он **шел** и куда **исчез**. Я только тогда **выпрямился** и **подумал**: «Зачем это отец ходит ночью по саду», — когда опять все **утихло** вокруг. Со страху я **уронил** нож в траву, но даже искать его не **стал**: мне **было** очень стыдно. Я разом **отрезвился**. Возвращаясь домой, я, однако, **подошел** к моей скамеечке под кустом бузины и **взглянул** на окошко Зинаидиной спальни. Небольшие, немного выгнутые стекла окошка тускло **синели** при слабом свете, падавшем с ночного*

неба. Вдруг — цвет их **стал изменяться**... За ними — я это **видел, видел** явственно — осторожно и тихо **спускалась** беловатая штора, **спустилась** до оконницы — и так и **осталась** неподвижной (И. С. Тургенев. Первая любовь).

Приведенный пример убедительно демонстрирует, сколь невелики фрагменты, удовлетворяющие правилу иконической интерпретации. Появление глаголов несовершенного вида может быть вызвано разными причинами: длительностью соответствующего действия, его повторяемостью, стремлением автора к наглядности (ср. последнее предложение). Оно также может маркировать переход от повествования к описанию (ср.: *Небольшие, немного выгнутые стекла окошка тускло **синели** при слабом свете, падавшем с ночного неба*) — подробнее об этом см. п. «Функционально-смысловые типы речи: описание, повествование, рассуждение».

Следует также подчеркнуть, что присутствие в тексте двух и более персонажей и попеременное упоминание об их действиях нарушает единство хронологии. В общем случае для каждого персонажа нужно строить свою шкалу времени. Применительно к приведенному выше отрывку действия рассказчика и его отца можно отразить при помощи двух соотнесенных между собой временных прямых, но в других случаях могут потребоваться гораздо более тонкий анализ и сложные построения. Некоторое представление об этом можно составить по следующему отрывку:

Он **надел** круглую шляпу *a la Van Dyck*, блузу, **взял** картон под мышку и **отправился**; я **попелся** вслед за ним. Ася **осталась** дома. Гагин, уходя, **попросил** ее позаботиться о том, чтобы суп был не слишком жидок; Ася **обещалась** побывать на кухне. Гагин **добрался** до знакомой уже мне долины, **присел** на камень и **начал срисовывать** старый дуплистый дуб с раскидистыми сучьями. Я **лег** на траву и **достал** книжку; но я двух страниц не прочел, а он только бумагу **измарал**... (И. С. Тургенев. Ася).

В приведенном фрагменте описываются различные (не коллективные) действия трех персонажей, причем порядок их упоминания в тексте применительно к Гагину и Асе не соответствует последовательности действий этих героев, а именно: Гагин *попросил* прежде, чем *отправился*, Ася *осталась* после того, как *обещалась*. Чтобы правильно выстроить порядок описанных здесь событий, лингвистический процессор должен отождествить действия, выраженные личным глаголом *отправился* и деепричастием *уходя*, связать во

времени уход Гагина, его просьбу и обещание Аси, а также поставить ее обещание раньше того момента, когда спутники оставили ее дома. Отсутствие в русском языке так называемого согласования времен затрудняет отслеживание ретроспекций, что может повлечь ошибки в интерпретации. Вот еще один показательный пример:

Я остановился, запыхавшись, на краю горы и, прислонясь к углу домика, стал рассматривать живописную окрестность, как вдруг слышу за собой знакомый голос:

— Печорин! давно ли здесь?

Оборачиваюсь: Грушницкий! Мы обнялись. Я познакомился с ним в действующем отряде. Он был ранен пулей в ногу и поехал на воды с неделю прежде меня (М. Ю. Лермонтов. Герой нашего времени).

Обратим внимание на последний абзац. Знания о мире диктуют читателю то очевидное и единственно верное прочтение, что знакомство Печорина с Грушницким, ранение последнего и его отъезд на воды произошли за какое-то время до того, как персонажи вновь встретились и обнялись. Однако применение принципа иконичности даст обратную (и нелепую) картину: персонажи обнимаются, еще не будучи знакомыми.

Иконичность в тексте не соблюдается также тогда, когда порядок перечисления событий регулируется не фактором времени, а какими-то другими соображениями. Возьмем такой (искусственно построенный) пример: *Иван достиг всего, чего хотел. Заработал кучу денег, удачно женился, построил дом на берегу моря, купил роскошную яхту...* Собственно говоря, по этому контексту нельзя сказать ничего определенного относительно временного порядка «достижений» Ивана, так как непонятно, какими соображениями руководствовался автор текста, выстраивая их в линейную цепочку: хронологией, иерархией ценностей (присущей самому автору, Ивану или адресату) или ходом своих воспоминаний. Суммируя вышесказанное, приходится констатировать, что правило иконической интерпретации статистически оправдано на небольших отрезках текста, но даже там не гарантирует стопроцентный успех.

Легко видеть, что оно не работает на более протяженных фрагментах, где повествование с высокой вероятностью нарушается различными отступлениями, ретроспекциями, рассуждениями, описаниями и т. д. Соответственно, к глаголам совершенного вида добавляются глаголы несовершенного вида (в широком диапазоне форм времени и наклонения), причастия, деепричастия; простые предло-

жения чередуются со сложными. Именно так организовано литературное повествование, а также жанры писем и мемуаров. Подтверждением могут служить даже весьма короткие отрывки текстов, ср.:

*По возвращении из деревни (19 декабря), куда он **ездил** на короткое время, Пушкин **приехал** прямо ко мне и **жил** в том же доме Ренкевича, который, как сказано, на Собачьей площадке **стоит** лицом, а задом **выходит** на Молчановку²⁴.*

*Однажды Ал. Серг-вич **получает** от Пуцина из Москвы письмо, в котором сей последний **извещает** его, что **едет** в Петербург и очень бы **желал увидеться** там с Ал. С-чем. Недолго **думая** пылкий поэт мигом **собрался** и **поскакал** в столицу. Недалеко от Михайловского, при самом почти выезде, **попался** ему на дороге поп, и Пушкин, **будучи суверен**, **сказал** при сем: «Не **будет** добра!» — и **вернулся** в свой мирный уединенный уголок. Это **было** в 1825 году, и провидению **угодно было осенить** своим покровом нашего поэта. Он **был спасен**!²⁵*

Обратимся теперь к отрезкам художественного повествования, построенных на употреблении глаголов несовершенного вида. Примечательно, что их последовательность также способна иконично отражать порядок действий, ср.:

*В девять часов из Парижа **выходит** специальный поезд, отвозящий в Гавр пассажиров «Нормандии». Поезд **идет** без остановок и через три часа **вкатывается** в здание гаврского морского вокзала. Пассажиры **выходят** на закрытый перрон, **подымаются** на верхний этаж вокзала по эскалатору, **проходят** несколько зал, **идут** по закрытым со всех сторон сходням и **оказываются** в большом вестибюле. Здесь они **сажаются** в лифты и **разъезжаются** по своим этажам (И. Ильф, Е. Петров. Одноэтажная Америка).*

В данном примере порядок следования глаголов несовершенного вида отражает этапы типовой процедуры. Однако чаще всего такие глаголы используются для придания большей наглядности передаваемым событиям, а также для различных отступлений (описаний, рассуждений и др.). В приводимом ниже отрывке, например, пере-

²⁴ Письмо С. А. Соболевского М. Н. Лонгинову // Вересаев В. В. Пушкин в жизни: Систематический свод подлинных свидетельств современников. Т. 1. М., 2001. С. 354.

²⁵ Лорер Н. И. Записи моего времени. Воспоминание о прошлом // Вересаев В. В. Пушкин в жизни: Систематический свод подлинных свидетельств современников. Т. 1. М., 2001. С. 308.

ключение с глаголов совершенного вида на глаголы несовершенного вида знаменует начало ретроспективного рассуждения, а обратный переход сигнализирует возврат к повествованию о событиях, ср.:

*Я назвал себя, и мы **разговорились**. Я узнал, что Гагин, путешествуя, так же как и я, для своего удовольствия, неделю тому назад **заехал** в городок Л., да и **застрял** в нем. Правду сказать, я неохотно **знакомился** с русскими за границей. Я их **узнавал** даже издали по их походке, покрою платья, а главное, по выраженью их лица. Самовольное и презрительное, часто повелительное, оно вдруг **сменялось** выражением осторожности и робости... Человек внезапно **настораживался** весь, глаз беспокойно **бегал**... «Батюшки мои! не соврал ли я, не смеются ли надо мною», — казалось, **говорил** этот уторопленный взгляд... **Проходило** мгновение — и снова **восстанавливалось** величие физиономии, изредка чередуясь с тупым недоумением. Да, я **избегал** русских, но Гагин мне **понравился** тотчас (И. С. Тургенев. Ася).*

Важно отметить, что наличие последовательности глаголов несовершенного вида в нарративном тексте (или отрывке) в общем случае не позволяет судить о его темпоральной организации. В качестве примера приведем два коротких фрагмента, принадлежащих перу одного и того же автора: в первом описываются одновременные события, во втором — последовательные, ср.:

*Однажды вечером я **был** дома, в своей одинокой комнате, и **занимался** тем, что **лежал** на диване, стараясь делать как можно меньше движений. Я человек прилежный, энергичный, и это занятие нисколько меня не **утомляло** (А. Аверченко. Мужчины).*

*Навстречу мне озабоченно и быстро **шагает** чиновник Хрякин, торопящийся на службу.*

*Увидев меня, он **расплывается** в изумленную улыбку (мы встречаемся с ним каждый день), быстро **сует** мне руку, **бросает** на ходу:*

— Как поживаете, что поделяваете?

*И **делает** движение устремиться дальше.*

*Но я **задерживаю** его руку в своей, **делаю** серьезное лицо и говорю... (А. Аверченко. На улице).*

Анализируя приведенные фрагменты, можно сделать предположение о том, что индикатором темпоральной структуры текста могут служить семантические признаки глаголов (точнее, их ССВ), а именно: поскольку в первом отрывке используются стативные глаголы, вероятно, соответствующие состояния симультанны, а на-

сыщенность второго фрагмента глаголами движения предполагает последовательность, сменяемость соответствующих действий. В дальнейшем будет показано, что в действительности все обстоит сложнее и семантические признаки глагола сами по себе еще ничего не говорят о временной структуре текста: одни и те же глагольные ССВ могут использоваться и в повествовании о событиях, и в описаниях. Однако в ряде случаев, где информация о признаках глагола подкрепляется семантикой наречий и союзов, данное предположение работает. Так, в нижеследующем примере глаголы чувственно-го восприятия и эмоционального состояния *прислушивалась*, *глядела*, *не хотела (вступать)* обозначают одновременные события, а словоформы *заговаривал* (начинательный способ действия), *поднималась* (смена позы), *удалялась*, *появлялась* (движение), *усаживалась* (смена позы) — последовательные. Временной порядок дополнительно подчеркивается союзом *как только*, наречиями *потом*, *опять*, ср.:

Она не то, чтобы избегала его... напротив, она постоянно садилась в небольшом от него расстоянии, прислушивалась к его речам, глядела на него; но она решительно не хотела вступать с ним в беседу, и как только он заговаривал с нею — тихонько поднималась с места и тихонько удалялась на несколько мгновений. Потом она появлялась опять, и опять усаживалась где-нибудь в уголку... (И. С. Тургенев. Вешние воды).

Еще одно соображение нуждается в том, чтобы на него обратить внимание. Для правильной интерпретации временного плана нарративного текста, помимо личных и неличных форм глагола, необходимо учитывать свернутые пропозиции: их также требуется локализовать во времени. Показателен следующий отрывок:

Болезнь остановила на время образ жизни, избранный мною. Я занемог гнилою горячкою. Лейтон за меня не отвечал. Семья моя была в отчаянии; но через шесть недель я выздоровел (А. С. Пушкин. Воспоминания).

Если не принимать во внимание предикатное существительное *болезнь*, получится, что определенный образ жизни героя остановился до того, как он занемог. На самом деле смысловые отношения между предикатами здесь не сводятся к отражению последовательности событий: предложение *Я занемог гнилою горячкою* служит для уточнения характера болезни. В рамках нашего подхода подобные случаи

автоматически отслеживаются и адекватно интерпретируются благодаря тому, что лексические единицы соотнесены с семантическими признаками вне зависимости от их частеречной принадлежности. В данном случае существительному *болезнь* в семантико-синтаксическом словаре приписано значение 'действие', и на этапе коммуникативно-синтаксического анализа оно трактуется как свернутая пропозиция.

В литературе неоднократно отмечалась сложность и многоаспектность темпоральной структуры художественного текста. Художественное (оно же — сюжетное, событийное) время отличается от реального времени по ряду параметров, а именно: оно может застыть, оно обратимо и хронологическая определенность для него несущественна [Тураева 2009: 23–29]. Поэтому привычно используемая для отражения природы реального времени бесконечная числовая ось (с присущими ей одномерностью, непрерывностью, односторонненностью) не является адекватным инструментом моделирования сюжетного времени. Само только присутствие в тексте более одного персонажа, как правило, предполагает несколько временных планов и, следовательно, требует введения соответствующего числа временных шкал.

Длительность художественного времени определяется длительностью событий фабулы. Сюжет нередко нарушает их хронологическую последовательность. Особенно наглядно это можно видеть на примере детективных произведений, где указанная особенность составляет неперемный атрибут жанра. Темпоральная организация детективного рассказа, а тем более романа отличается высокой сложностью. Фабула является как бы двусоставной, включая линию преступления и линию расследования; в основной части повествования эти линии представлены фрагментарно и непоследовательно, и лишь в конце произведения, как правило, сыщик (в присутствии заинтересованных лиц) осуществляет полную реконструкцию случившегося, т. е. фабула разворачивается в хронологическом порядке (подробнее см. [Скребцова 2012]).

Филологический анализ выявляет присутствие в художественных произведениях несколько типов времени: авторского, событийного, субъективного времени рассказчика и моделируемого условного времени читателя [Николина 2004: 169–170]. Скрупулезно анализируя соотношения между различными видовременными формами русских глаголов и создаваемыми темпоральными планами текста, Н. А. Николина получает в итоге громоздкую систему,

которую, выражаясь языком логики, можно охарактеризовать как много-многозначное соответствие. Излишне говорить, что этот тип связей представляет наибольшие сложности для моделирования в системах искусственного интеллекта.

Руководствуясь суждением М. М. Бахтина о том, что «жанр и жанровая разновидность определяется... хронотопом, причем в литературе ведущим началом в хронотопе является время» [Бахтин 1975: 235], Н. А. Николина пишет, что каждый жанр отличается своим особым соотношением видовременных форм, доминирующей ролью отдельных форм времени. Автор иллюстрирует этот тезис примерами, утверждая, что для афоризмов и сентенций характерно употребление форм настоящего абстрактного или настоящего постоянного, для дневников — взаимодействие глагольных форм в прошедшем перфектном с формами настоящего актуального, для поэтической элегии — сопоставление форм прошедшего и настоящего или будущего, а также финал, подчеркивающий «точечность» сюжета, исчерпанность ситуации или ее неожиданное разрешение, для жанра автобиографии — сопоставление двух временных планов, открытость финала, субъективная сегментация времени и пр. [Николина 2004: 230–262, 267]. Мы намеренно позволили себе это достаточно длинное изложение результатов исследования, чтобы наглядно показать их непригодность, бесполезность для решения задачи автоматической обработки текста. Наше замечание ни в коей мере не оспаривает ценности цитируемой монографии; оно лишь призвано акцентировать тот факт, что имеющиеся описания в подавляющем большинстве имеют филологическую направленность и мало приспособлены к нуждам компьютерной лингвистики.

В нашей практической работе мы сознательно следовали определенным ограничениям. Так, мы не учитывали причастия и деепричастия с точки зрения их локализации на темпоральной оси по отношению к личным формам глагола. Мы ограничились анализом монологических фрагментов, поскольку прямая речь персонажей ставит на повестку дня вопросы, связанные не только с фактуальным содержанием текста (диктумом), но и с его оценкой (модусом), а это выводит нас за рамки текущей темы. Мы также не касались вопроса о сфере действия известного схоластического принципа *post hoc ergo propter hoc*, а именно: в каких случаях при моделировании информационной структуры текста можно ограничиться собственно темпоральными отношениями между событиями, а в каких

требуется добавлять к ним причинно-следственную связь?²⁶ Здесь, как и во многих других затронутых выше аспектах, дело упирается в разработку энциклопедического компонента, описывающего внешний мир, его восприятие человеком и взаимодействие людей с миром и между собой.

Как уже отмечалось, серьезным препятствием на пути автоматической интерпретации темпоральной организации текста является высокий индекс грамматической неоднозначности глагольных форм в русском языке, наличие не только центральных (прямых) употреблений временных форм глаголов, но и многочисленных периферийных (переносных). Так, глагол в настоящем времени может описывать события в прошлом («настоящее историческое») или обозначать будущие действия — намеченные или воображаемые. Формы прошедшего времени могут использоваться для обозначения событий в будущем или в контексте абстрактного настоящего. Формы будущего времени в переносных употреблениях могут отсылать к настоящему (неактуальному) или прошлому²⁷.

Из пяти видовременных форм разве что глаголы несовершенного вида в прошедшем времени практически всегда используются «по назначению», что, впрочем, все равно оставляет простор для некоторой неопределенности (например, по параметру единичности/повторяемости действия, ср. *приходил, спрашивал* и т. п.). Остальные четыре формы «приписаны» к тому или иному временному плану в соответствии со своей прямой функцией, что не мешает каждой из них отсылать к двум другим фрагментам оси времени при переносных употреблениях. Следует заметить, что различные литературные жанры как раз и характеризуются широким диапазоном прямых и переносных употреблений временных форм глаголов. Поэтому с точки зрения темпоральной организации художественное повествование является, по-видимому, наиболее сложным для автоматической обработки типом текста.

²⁶ Ответ на этот вопрос, в свою очередь, обуславливает выбор в пользу событийной или пропозитивной (фактивной) интерпретации. Ср. предложение *С приходом к власти военной хунты в стране начался упадок культуры*, которое допускает как событийное прочтение ('После того как к власти пришла военная хунта...'), так и фактивное ('Вследствие того, что к власти пришла военная хунта') [Арутюнова 1988: 185].

²⁷ Иллюстрации см. в [Бондарко 1971а; Русская грамматика 1980, т. 1: 628–636; Падучева 2010].

Рассмотрим теперь, как в этом плане устроены другие стили речи — научный и публицистический.

В научных текстах могут встречаться все пять видовременных форм глагола, обозначающих события, которые относятся к прошлому, настоящему или будущему. Как правило, эти формы употребляются здесь в своих прямых функциях. Например, жанр отчета о прошедшем научном событии характеризуется преимущественным употреблением форм прошедшего времени и исходной привязкой к абсолютной шкале отсчета. Вот характерный пример²⁸:

24 января 2013 г. в Институте русского языка им. В. В. Виноградова РАН (Москва) **прошли** очередные (сорок четвертые) Виноградовские чтения. В этом году они **были посвящены** проблемам изучения истории языка. Организаторы чтений **приняли** решение открыть заседание докладом, посвященным вопросам этимологии.

В докладе Ж. Ж. Варбот (Москва) «О соотношении структурно-семантического аспектов анализа в праславянской этимологии» **рассматривается** «вечный» для этимологической методики (поставленный еще в конце XIX в.) вопрос о соотношении (связи/приоритете) одного из двух основных аспектов этимологического анализа: структурного и семантического. Докладчик **предлагает** различать обращение к этим двум аспектам на уровне оценки этимологического решения и на уровне стимула к разработке нового решения. <...>

В докладе А. М. Молдована (Москва) «К текстологии 16 Слов Григория Богослова с толкованиями Никиты Ираклийского» **было показано**, что большая часть текстологических особенностей списков древнейшей редакции Слов появилась позднее, когда уже были переведены Толкования <...>

С. М. Михеев (Москва) в докладе «О древнерусских глаголических надписях» **рассказал** о выявленных в настоящее время глаголических надписях-граффити на территории Древней Руси <...>

С. М. Толстая (Москва) в сообщении «Из истории понятий: “честь” и “честность”» **отмечает**, что концепт “честь” имеет две стороны — “внешнюю” (почитание, уважение и далее — почет, почести, знаки уважения, награда) и “внутреннюю” (высокие моральные качества,

²⁸ Капитанова Ю. С. Виноградовские чтения 2013 г. // Вопросы языкознания. 2013. № 5. С. 156–157.

моральный кодекс). В русском языке оба понятия **выражаются** словом “честь” и его дериватами <...>

В докладе А. А. Пичхадзе «О взаиморасположении глагольной формы и прямого дополнения в придаточных предложениях в древнерусском языке» **шла речь** о порядке в группе “личная форма глагола + прямое объектное дополнение” в независимых и — подробно — зависимых предикациях. <...>

В докладе П. В. Петрухина (Москва) **рассматривались** древнерусские двойные глаголы — бессоюзные сочетания двух глаголов, обладающие рядом особых семантических, синтаксических и просодических свойств <...>

М. А. Малыгина (Москва) в сообщении «Особенности восприятия греческого текста славянскими переводчиками» **рассмотрела** некоторые ошибки перевода с греческого языка в типографических текстах <...>.

Напротив, анонс (о грядущем научном событии), сообщая о том, что состоится в будущем, тем не менее допускает употребление форм настоящего времени у глаголов определенных семантических групп (ср. *планируется, принимаются* и т. д.), ср.:

8–11 мая 2011 г. в Шанхае (Китай) **состоится** XII Конгресс МАПРЯЛ «Русский язык и литература во времени и пространстве». На конгрессе **предполагается** работа по следующим направлениям: государственная языковая политика; роль и место русского языка в государстве и межгосударственном общении; «Русский мир»: русская культура, русский менталитет в условиях современной глобализации; новое в системно-структурном описании современного русского языка; русский язык: диахрония и динамика языковых процессов и др. Заявки на участие (см. ниже) и тезисы объемом до 1 страницы **принимаются** до 31 мая по адресу <...>²⁹

В обоих этих стилевых жанрах правило иконической интерпретации не действует, даже если имеет место последовательность грамматически однородных глагольных форм. В отличие от жанра коротких устных рассказов [Labov, Waletzky 1967], порядок следования предложений определяется здесь не временным фактором, а достаточно жестким шаблоном. Это лишний пример того, что успешная интерпретация текста невозможна без учета типа текста.

²⁹ Respectus Philologicus. 2010. № 17. С. 249.

Хорошо структурированным является также жанр аннотации к научной статье или монографии. С точки зрения темпоральной организации, аннотации обычно представляют собой набор утверждений в настоящем времени, касающихся цели, материала, значимости и основных результатов исследования. Временные рамки соответствующих пропозиций «слева» (на оси времени) ограничены моментом порождения текста, а «справа» не имеют границ. Как и в рассмотренных жанрах отчета о прошедшей конференции и ее анонса, все предикаты здесь являются симультанными: обработка каждого последующего ПСДО приводит к добавлению новых объектов и признаков, не изменяя и не отменяя ранее созданные.

Что касается жанра научной статьи, он сам по себе неоднороден. Так, Е. С. Троянская [1989: 24] выделяет, в зависимости от обсуждаемого предмета и целей авторов, следующие разновидности: 1) статья — краткое сообщение о результатах научно-исследовательских и опытно-конструкторских работ; 2) собственно научная (научно-техническая) статья; 3) передовая статья; 4) историко-научная статья; 5) дискуссионная / полемическая статья; 6) научно-публицистическая статья; 7) рекламная статья.

Т. В. Радзиевская [1986] проводит иное деление, противопоставляя теоретические научные тексты, в основе которых лежит рассуждение, эмпирическим текстам, сводящимся к описанию. Можно утверждать, что последний тип текстов отличается большей однородностью: научные статьи (или их фрагменты), целью которых является описание объекта исследования, строятся преимущественно в настоящем времени. Это же характерно и для соответствующего типа научных монографий, а также недидактических фрагментов учебных пособий. Здесь также все признаки являются симультанными, так что построение информационной структуры данных по тексту сводится в основном к добавлению новых объектов и связей, ср.:

*Если психолога **интересуют** реакции индивида на общественные структуры, а социолог **пытается объяснить** те же самые факты в терминах общества, то для антрополога основной интерес — это ценности социокультурной группы и культурные правила поведения, которые **раскрывают** эти ценности. Иначе говоря, для антрополога каждый акт языкового выбора **представляется** значимым как средство выражения — и постижения — ценностей данной культуры.*

*Антропологический подход, кроме того, **отличается** методикой исследования. Основная цель — изучить нормальное, неконтролируе-*

мое поведение; основной антропологический метод — включенное наблюдение (*participant observation*). Антрополог **включается** в жизнь общины, чтобы иметь возможность наблюдать жизнь и поведение людей в нормальных условиях³⁰.

Прошедшее время в научном тексте используется при обращении к истории предмета или его изучения; при этом возможное чередование с формами настоящего времени не создает композиционного разрыва. Иллюстрируя этот тезис, Д. Н. Шмелев [1977: 139] приводит фрагмент из книги А. А. Шахматова «Очерк современного русского литературного языка». Мы ограничимся отдельной фразой из него, ср.: «Другой процесс *лежит* в основании образования романских языков: вульгарная латынь *усваивалась* народными массами (преимущественно кельтами) и *вытесняла* их, исконно им свойственный язык». Это существенная особенность научного стиля речи по сравнению с художественным [Шмелев 1977].

В теоретических работах, направленных на защиту определенных положений и критику чужих точек зрения, как настоящее, так и прошедшее время могут употребляться при изложении позиций сторон. К примеру, допустимы оба варианта: В «*Логико-философском трактате*» Людвиг Витгенштейн *утверждал/утверждает...* Теоретическая научная статья по своей стилистике сближается с публицистическими текстами, для которых актуальной является задача автоматического разделения фактов и мнений (подробнее см. ниже).

Специфическими чертами научной статьи являются выраженная метаязыковая функция (присущая также учебным текстам) и, как следствие, присутствие так называемого метатекста [Вежбицка 1978] или дейксиса дискурса [Макаров 2003: 198], регулярно реализующегося в конструкциях типа *как мы подчеркивали выше, как будет показано ниже, из сказанного следует, в дальнейшем мы покажем, рассмотрим теперь, см. далее* и т. п. Вследствие этого текст научной статьи оказывается неоднородным: утверждения, направленные на изложение содержания исследования (цель, предмет, процедура и пр.), перемежаются высказываниями, регулирующими способ изложения. Темпоральная организация повествования получается как бы двухслойной: на «обычное» (внешнее) время, связанное с изложением содержания исследования и выражением мнений,

³⁰ Вахтин Н. Б., Головкин Е. В. Социоллингвистика и социология языка: учеб. пособие. СПб., 2004. С. 105.

«наслаивается» внутреннее время повествования³¹ (ср.: *начнем с..., покажем теперь, как следует из примеров выше, в заключение подчеркнем...*). Соответственно, возникает задача дифференциации этих двух типов высказываний, каждый из которых может быть представлен широким диапазоном глагольных форм.

Достаточно надежными маркерами использования внутреннего времени служат лексические единицы *выше, ниже, здесь, далее* и некоторые другие в составе сложных слов или в сочетании с глаголами речи, аргументации, демонстрации. Очевидно, однако, что их недостаточно, чтобы выделить все высказывания, связанные с регуляцией способа изложения. Другая группа слов, которые могут служить данной цели, — это формы совершенного вида будущего времени глаголов указанных лексико-семантических групп (типа *рассмотрим, обратимся, перейдем, заметим* и т. п.). Помимо этого, имеет смысл отслеживать употребление прилагательных типа *предыдущий, последующий, дальнейший*, которые, как правило, в научных публикациях употребляются дейктически, хотя случаются и отклонения от этого правила.

Что касается упоминания в тексте научной статьи об авторах (*я, мы, мой, наш, автор*), оно может встречаться как в контексте внутреннего, так и внешнего времени, ср.: *Выше я (мы) указывал на то, что...* (внутреннее время) vs. *В своих предыдущих публикациях я (мы) неоднократно подчеркивал(и), что...* (внешнее время). Более того, некоторые контексты могут оказаться неоднозначными в рассматриваемом отношении, ср.: *Я уже указывал на то, что...* Снятие такой неоднозначности может быть достигнуто при помощи учета графематической информации³²: если предложение сопровождается сноской на научную работу, автор которой совпадает с автором текущей статьи, мы имеем дело с внешним временем, а если сноска отсутствует — с внутренним. Разумеется, разработка подобного алгоритма требует дополнительных усилий, но в принципе осуществима.

* * *

Публицистический стиль охватывает большое число разнообразных речевых жанров. С точки зрения временной организа-

³¹ Оно же — «время адресата», «время читателя» [Николина 2004: 167; 170].

³² Напомним, что результаты обработки текста по мере движения от одного этапа к другому (графематика, морфология, семантика, синтаксис) не теряются: соответствующая информация накапливается и в любой момент может быть использована во всем объеме.

ции газетные тексты могут оказаться не менее сложными, чем художественные. Если в художественной литературе проблемы, возникающие при автоматической обработке, в значительной степени обусловлены частыми переходами от повествования к описанию и обратно, в текстах СМИ существенным осложняющим фактором является присутствие аргументативного начала. Практически все публицистические жанры (за исключением разве что хроникальной и короткой информационных заметок), наряду с рассказом о событиях, подразумевают более или менее эксплицитное выражение различных мнений, принадлежащих как третьим лицам (названным или подразумеваемым), так и самому автору. Пропозиции в тексте могут относиться как к событийному содержанию текста, так и к его оценочному компоненту.

Здесь мы сталкиваемся со сложной проблемой автоматического разграничения фактов и мнений. Информационная структура должна включать оба эти компонента, но учитывать принципиальное различие между соответствующими типами данных. В описываемом лингвистическом процессоре акцент сделан на событийном компоненте, а оценки прикрепляются к соответствующим пропозициям. Формат представления мнений в настоящее время находится в стадии разработки. Предполагается некоторым образом их систематизировать посредством введения шкалы или закрытого списка значений, что, конечно, приведет к известному огрублению, но зато позволит эффективно обрабатывать большие массивы текстов. Помимо этого, оценочный компонент должен включать ссылку на автора и тем или иным способом локализоваться на шкале времени.

В связи с недостаточной разработанностью данной проблемы обратимся к рассмотрению того жанра публицистического стиля, который лишен (эксплицитной) оценочности, а именно короткой информационной заметки. Несколько упрощая положение вещей, будем считать, что в ней вовсе нет оценочного компонента, а содержание исчерпывается изложением конкретных событий.

Отличительными особенностями данного жанра являются сильно выраженная привязка к моменту создания и опора на абсолютную систему отсчета времени. В дискурсивных исследованиях отмечается, что новостные тексты СМИ структурированы по принципу релевантности, т. е. наиболее важная информация представлена в начале сообщения [Ван Дейк 1989: 132–133]. В связи с постоянным обновлением содержания газетных новостей их темпоральная организация демонстрирует что-то наподобие обратного хронологиче-

ского порядка повествования: сначала идет самая свежая информация, отражающая события на момент создания репортажа, а затем предыстория (фон), разворачивающаяся назад, ко все более ранним стадиям развития сюжета. В литературе это иногда называется «структурой перевернутой пирамиды» (см., например, [Bell 1998]).

Заметим, однако, что приведенные суждения относятся к основной части информационной заметки. Что касается рамки, в которую она заключена, первое предложение, как и заголовок (если он есть), обычно содержит формулировку темы всего сообщения, в то время как последнее (или последние) возвращает читателя в настоящее (момент создания сообщения). Получается, что в целом этот жанр похож на бытовой нарратив, который изучал У.Лабов (см. выше в данном параграфе), за тем исключением, что в основной части можно наблюдать действие «обратного иконического принципа». Покажем это на следующем примере:

***На Васильевском острове угнали машину
Управления делами президента России***

На Васильевском острове угнали машину Управления делами президента России.

*Как передает АН «Оперативное прикрытие», 2 мая в 5:20 утра в районную полицию **обратился** заместитель директора ФГУП «Кремлевский». По словам 43-летнего чиновника, около 1:30 он **обнаружил** пропажу служебного автомобиля BMW-740 с московскими номерами. Иномарка **была припаркована** у дома 30 по улице Кораблестроителей.*

Размер ущерба устанавливается. Машина оборудована штатной сигнализацией, которая была включена³³.

Однако не следует возводить структуру «перевернутой пирамиды» в ранг универсального правила. Она действительно работает, в случае если новостной текст повествует о заметном происшествии, его возможных причинах и последствиях. Вместе с тем жанр короткой информационной заметки, выделяемый по формальному критерию «10–30 строк» [Лысакова 2005: 115, 119], в содержательном плане является неоднородным. Другой вид информации подразумевает иные логико-смысловые отношения между предложениями и иную темпоральную структуру текста. Проиллюстрируем наше суждение заметкой, включающей выражение мнения и его обоснование, ср.:

³³ URL: <http://www.gazeta.spb.ru/1634263-0/> (дата обращения: 26.03.2018).

Эксперт: Российские товары не смогут заменить импортные

Прилавки магазинов в результате запрета на импорт товаров из Евросоюза скорее опустеют, чем «обрусеют», считает ритейлер, управляющий партнер Management Development Group Дмитрий Потапенко. Об этом он заявил в интервью в студии телеканала РБК.

«Я с большим восторгом слушаю отечественных производителей, когда их спрашивают журналисты или люди с улицы: у вас отечественный товар? Он говорит: да, отечественный. Я говорю: хорошо, давайте табличку разложим. Оборудование чье? Импортное — немецкое, голландское. Семенной материал чей? Голландский, канадский, и далее по списку. Для того, чтобы говорить “отечественный товар”, нужно сначала лет двадцать возвращать ученых, которые займутся семенным материалом, вложить в них кучу денег, они будут долго ошибаться, и только после этого у нас появится что-то свое, — заявил эксперт. — Что мешало последние двадцать лет поддерживать нашего сельхозпроизводителя?»³⁴.

Нетрудно видеть, что в данном тексте отсутствует структурирование по принципу «перевернутой пирамиды». Другая разновидность газетного сообщения, в котором также не наблюдается иконической упорядоченности (ни прямой, ни обратной), представлена текстами, где общий тезис, вынесенный в заголовок, и первое предложение далее конкретизируются частными утверждениями о фактах, подкрепляемыми цифрами, ср.:

В некоторых регионах РФ из-за санкций дорожают продукты

После введения Россией эмбарго на поставки продуктов из ЕС и США в стране начали расти оптовые цены на некоторые товары в Москве и регионах. Об этом пишет газета «Коммерсантъ».

Так, по данным издания, в российской столице свинина для мясоперерабатывающих комбинатов подорожала на 6 %. В Приморье некоторые виды рыбы подорожали на 40 %, а в Сахалинской области рост цен на куриные окорочки достиг 60 %.

Об этом на заседании совета по продовольственной безопасности правительства Сахалинской области заявил региональный министр сельского хозяйства, промышленности и торговли Николай Борисов.

³⁴ URL: <http://www.rosbalt.ru/video/2014/08/07/1301485.html> (дата обращения: 26.03.2018).

По его словам, в Южно-Сахалинске и близлежащих городах оптовые цены, к примеру, на сыры выросли на 10 %, а на мясо — на 15 %.

В Приморье мясо подорожало на 26 %, импортируемые из Китая яблоки — почти на треть. На острове розничные цены на продовольствие увеличились на 3–60 %, поясняют представители торговых сетей Сахалина³⁵.

Итак, несколько неожиданно, жанр короткой информационной заметки оказался внутренне неоднородным с точки зрения типов содержательных связей между заголовком и основным текстом, а также способов развертывания темы. Отсюда следует невозможность постулирования общей для данного жанра временной структуры текста.

* * *

В рамках данного раздела мы стремились показать разнообразие темпоральной организации в текстах различных жанров и неоднородность видовременных форм глаголов в русском языке. Цель этого аспекта анализа предложения заключается в том, чтобы вскрыть и описать те формальные признаки, которые можно использовать при построении алгоритма автоматического определения временной структуры текста. Речевые жанры широко варьируют по степени схематичности, и только очень немногие укладываются в определенный ясный шаблон³⁶. Представляется, что рассмотренный материал (примеры литературного, научного и публицистического повествования, прескриптивных и дескриптивных текстов) убедительно демонстрирует сложность задачи.

АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ ТИПА ТЕКСТА

Понятие типа текста

В предыдущем параграфе мы сделали попытку охарактеризовать хотя бы в грубом приближении особенности временной организации текстов отдельных жанров и соотнести ее с интерпретацией грамматических форм глагола. Следует признать, что в наших рассуждениях было некоторое лукавство: ведь тем самым предполагалось, что жанр обрабатываемого текста заранее известен. Вместе с тем в пре-

³⁵ URL: <http://www.24rus.ru/more.php?UID=113839> (дата обращения: 26.03.2018).

³⁶ К шаблонам относятся прежде всего жанры официально-делового стиля, оформление которых регулируется нормативными документами.

дисловии к настоящей монографии мы подчеркивали, что лингвистический процессор не знает, с каким текстом (стиль, подстиль, жанр) он работает. Поэтому было бы некорректно обойти стороной вопрос о том, каким образом возможно осуществлять автоматическое распознавание типа текста. Как и в случае с определением темпоральной структуры, речь идет о различных видах информации, приписанной на более ранних этапах анализа, которая может служить своего рода «опознавательными знаками» того или иного типа текста.

Прежде чем обсуждать эту тему, следует пояснить содержание используемого здесь понятия «тип текста». Как будет видно из дальнейшего изложения, в зависимости от ситуации оно может соответствовать функциональному стилю, подстилю, отдельному жанру или даже функционально-смысловому типу речи в рамках жанра (ср. чередование повествования и описания в художественном произведении, описания и рассуждения в научной или публицистической статье и т. д.). Предвосхищая упреки, связанные с широтой и размытостью введенного понятия, мы апеллируем к соображениям практического удобства, важным в контексте прикладного исследования. Выражение «тип текста» позволяет избежать обсуждения широкого спектра проблем, связанных с классификацией текстов по различным основаниям (общее представление об этом можно составить, например, по работам [Beaugrande 1980: 197–199; Богданов 1993: 9–17; Филиппов 2003: 183–205], номенклатурой функциональных стилей, подстилей и жанров русской речи, соотношением понятий стиля и жанра (см., например, [Орлова 1997]), набором конституирующих признаков речевых жанров, их обусловленностью коммуникативной ситуацией и пр. Эти сложные и спорные вопросы, относящиеся к теории функциональной стилистики, на протяжении многих десятилетий дискутируются в соответствующей литературе (см., например, [Жанры речи 1997; 1999; 2002]), однако однозначного решения до сих пор не получили, и едва ли оно возможно в принципе. Наша задача, напротив, вполне конкретная, и заключается она, как уже было сказано, в том, чтобы постараться обнаружить достаточно надежные дифференциальные признаки однородных в стилистическом отношении текстов, с тем чтобы обеспечить их успешную автоматическую обработку³⁷.

³⁷ Ср.: «Понятие “тип текста” в настоящее время принято как рабочий термин в современных исследованиях по теории текста, в частности в лингвистике текста. Обозначает он эмпирически существующие формы манифестации текстов.

Сама идея поиска отличительных признаков того или иного стиля (жанра) не является новой для языкознания, в том числе отечественного. Начиная с 1960-х годов в нашей стране проводились соответствующие исследования, правда, многие из них имели выраженную филологическую или учебно-методическую ориентацию, так что результаты не представляют большой ценности для компьютерной обработки языка. В это же время, однако, наблюдался рост интереса к применению количественных методов в языкознании и складывались различные отрасли лингвистической статистики, в том числе стилеметрия, основной (и исторически первичной) задачей которой является атрибуция текстов. Последняя, как правило, предполагает установление авторской принадлежности произведения, когда оно анонимно, подписано псевдонимом или в случае литературных мистификаций (ср. [Марусенко 1990]). Вместе с тем существует и более широкое понимание атрибуции, ср.: «Наряду с авторской выделяют не авторскую атрибуцию, цель которой — отнесение данного текста к определенному языку, стилю, периоду времени, литературной школе, литературному направлению, идейному течению, общественному классу и т. д.» [Мартыненко 1988: 6].

В области квантитативной типологии текстов, становление и расцвет которой связан с периодом 1960–1980 гг., были получены некоторые результаты, касающиеся формальных критериев дифференциации различных типов текстов, таких как средняя длина слова, предложения, относительная частота словоупотреблений по частям речи и пр. [Статистичні параметри... 1967; Статистика речи... 1973; Алексеев 1983; Лингвистические распределения... 1985]. Следует заметить, что в рамках данной задачи (в отличие от предшествующих) мы допускаем использование количественных методов. Это связано с тем принципиальным обстоятельством, что текст не является очередным (следующим за синтаксическим) уровнем языковой структуры³⁸. Его построение регулируется не только правилами

Расхождения в толковании понятия “тип текста” еще достаточно велики. Оно трактуется то слишком узко, то слишком широко...» [Валгина 2003: 112].

³⁸ Ср. знаменитые слова Э. Бенвениста: «...нужно признать, что категорематический уровень включает только одну форму языковых единиц — предложение. Оно не составляет класса различимых единиц, а поэтому не может входить составной частью в единицу более высокого уровня. Предложение может только предшествовать какому-нибудь другому предложению или следовать за ним, находясь с ними в отношении последовательности. Группа предложений не образует единицы высшего уровня по отношению к предложению. Языкового уровня,

языка, но и внешними, экстралингвистическими факторами³⁹, следовательно, собственно лингвистического анализа может оказаться недостаточно.

Рассуждая об отличительных признаках стиля, Д. Н. Шмелев [1977: 80] пишет, что «различные функционально-стилистические разновидности совсем не в одинаковой степени характеризуются именно особыми, присущими только им языковыми средствами. Одни из выделяемых разновидностей характеризуются использованием по существу всех средств языка, другие в большей или меньшей степени ограничены в этом отношении». Самым замкнутым функциональным стилем является официально-деловой стиль, на втором месте — научный стиль. Публицистика, по словам автора, отличается нестабильным репертуаром «собственной» лексики и фразеологии и почти неограниченной возможностью привлечения самых разнообразных языковых средств. Применительно к языку художественной литературы едва ли можно говорить о наличии каких-либо специфических языковых средств [Шмелев 1977: 80-81].

Из сказанного следует, что и возможности распознавания стилей различны: чем он более однороден, тем легче его узнать и наоборот. Сложнее всего обстоит дело с идентификацией художественного стиля; скорее всего, ее следует проводить «по остаточному принципу». Но говоря в целом, следует признать, что различия, как правило, не столь существенны, чтобы на них можно было уверенно опираться при определении типа текста. Дело осложняется и тем, что речевые жанры, их структурные и языковые особенности исследованы фрагментарно, несистемно и явно недостаточно для того, чтобы можно было воспользоваться готовыми описаниями. Требуется огромная предварительная работа по анализу разнообразных типов текстов и выработке соответствующего процедурного представления.

Обратимся теперь к краткому обзору имеющихся в литературе наблюдений, касающихся различных типов текстов в интересующем нас аспекте. Руководствуясь суждением М. М. Бахтина [1986] о том, что каждый жанр характеризуется особой тематикой, композицией и стилистикой, постараемся обнаружить соответствующие призна-

расположенного выше категорематического уровня, не существует» [Бенвенист 1974: 139].

³⁹ Это обстоятельство усиленно подчеркивали видные представители дискурсивного анализа (см., например, [Coulthard 1977: 4, 7; Stubbs 1983: 102–103]), стремясь преодолеть заблуждения, распространенные на ранней стадии развития лингвистики текста.

ки для разного рода текстов, при этом делая основной упор на выявление именно языковых (лексических, морфологических, синтаксических) различий.

Научный стиль

Наиболее обстоятельно в литературе описан научный стиль речи, чьи структурные и языковые особенности получили отражение в ряде монографий и коллективных сборников (см., например, [Язык научной литературы 1975; Особенности... 1976; Язык и стиль... 1983; Кожина 1966; 1972; 1986; Лариохина 1979; Лингвистические особенности... 1981; Ванников 1984; Научная литература... 1985; Разновидности... 1989]), не говоря уже о множестве статей. Имеются работы, посвященные конкретным жанрам научной речи — научному реферату [Паташюте 1977; Закономерности... 1982; Грязнухина 1985; Радзиевская 1986], рецензии [Синдеева 1982; Кудасова 1983; Ляпунова, Мешков, Терехова 1990; Красильникова 1999], обзору [Воробьева 1986], статье [Клейменкова 1989; Вагурска 1992], аннотации [Яцко 1993]⁴⁰. Хотя большинство работ датируется еще доперестроечным периодом, основные характеристики жанров сохраняются и в наши дни, что объясняется высокой консервативностью научного стиля (по сравнению, скажем, с публицистическим).

Среди общих черт научного стиля исследователи отмечают почти исключительное использование формы 3-го лица глагола, заметное преобладание форм настоящего времени, сравнительно высокую долю существительных среднего рода, явный перевес родительного падежа (и прежде всего родительного приименного) над прочими падежными формами; в синтаксическом плане — широкое использование пассивных и безличных конструкций, предложений с составным именным сказуемым, сложных предложений союзного типа [Кожин, Крылова, Одинцов 1982: 94–99].

Широкомасштабное исследование [Кожина 1966] было направлено на то, чтобы при помощи скрупулезных статистических подсчетов выявить специфику научного текста по сравнению с художественным (из всех стилей наиболее далеко от него отстоящим). Однако полученные М. Н. Кожинной количественные различия в ос-

⁴⁰ Приводимые перечни, разумеется, не претендуют на полноту, в частности не упоминается обширная учебно-методическая литература, посвященная выработке навыков создания первичных и вторичных документов научного стиля речи на родном и иностранном языках.

новном не превышают нескольких процентов, а потому едва ли могут служить достаточным основанием для разграничения этих стилей (ср. [Комарова 2004: 161]). Причина тому — высокая неоднородность научного стиля речи, включающего большое число разнообразных жанров: письменных и устных⁴¹, монологических и диалогических, первичных и вторичных, обладающих открытой или, напротив, жестко фиксированной структурой.

Помимо собственно научных жанров, образующих ядро данного стиля (монография, статья, диссертация, доклад и пр.), выделяются также справочно-энциклопедические, научно-оценочные, научно-учебные, научно-деловые и другие типы текстов, относящиеся к периферийным и пограничным областям [Троянская 1989]. Более того, внутри отдельного жанра (например, научной статьи) могут проводиться дальнейшие подразделения и выделяться частные разновидности [Троянская 1989: 24]. Наиболее ярко черты научного стиля выражены в ядерных жанрах; остальные несут на себе отдельные характеристики смежных стилей. Так, в жанре научной рецензии, близком к публицистическому стилю, сильно выражено личностное начало и широко используется оценочная лексика [Красильникова 1999]. Научно-деловые жанры (описание изобретения, патент, авторское свидетельство) несут на себе черты официально-делового стиля, что проявляется прежде всего в их синтаксических особенностях.

С учетом вышесказанного можно предположить, что анализ однородного в жанровом отношении материала способен дать весьма убедительные результаты. Действительно, в коллективной монографии [Закономерности...1982], посвященной статистическому анализу текстов научных рефератов, в частности, содержатся следующие впечатляющие утверждения: «...из 20 возможных в системе спряжения глагольных синтетических форм — инфинитив, шесть личных форм настоящего времени, шесть личных форм будущего времени, четыре формы прошедшего времени, три формы повелительного наклонения — употребляются (с относительной частотой не менее 0,40 % каждая) лишь восемь, т.е. 40 %. Их суммарная частота составляет 99,9 % всех употреблений глагольных словоформ. При этом только три глагольные формы — инфинитив, 3-е лицо единственного и множественного числа настоящего времени — охватывают около 97 % всех употреблений глагола, в том числе фор-

⁴¹ Напомним, что устная речь не рассматривается в рамках данной работы.

мы настоящего времени 77,23 %...». Ценность подобной статистики в контексте автоматического распознавания типа текста бесспорна. Правда, далеко не все жанры характеризуются наличием столь устойчивых инвариантных признаков; к тому же данное исследование является заметным исключением в широком ряду работ по функциональной стилистике.

Если даже опора на грамматическую информацию не позволяет уверенно распознавать научную речь, тем более это справедливо в отношении лексики. Достаточно надежным показателем, по-видимому, может служить только высокое содержание терминологических единиц, относящихся к соответствующей дисциплине, ср.: «Ведущим признаком этого (научного — *Авт.*) стиля нужно признать терминологичность словарного состава» [Гальперин 1965: 69]. В то же время это явление может наблюдаться и в публицистике, ср. аналитические статьи по экономике и финансам в деловых изданиях. Что же касается так называемой общенаучной лексики (*метод, критерий, структура, анализировать* и т. п.), ее присутствие в тексте тем более не гарантирует его принадлежности к рассматриваемому стилю.

Вообще, как показано в исследовании А. И. Комаровой [2004], неоднородность, «разножанровость» научной речи влечет наличие, наряду с инвариантными, вариативных признаков регистра. Допустимыми становятся даже «стилистически маркированные научные тексты», содержащие идиоматические выражения, метафорические употребления слов, экспрессивные лексические единицы. В целом эти черты не свойственны научной речи, но полностью исключить их появление в тексте нельзя.

Представляет интерес замечание Л. В. Красильниковой [1999: 14] о невозможности употребления в научной речи местоимений *ты* и *вы* и, соответственно, глагольных форм второго лица. Казалось бы, это наблюдение могло бы быть успешно использовано для автоматического отграничения научного стиля от художественного и публицистического. В то же время следует заметить, что оно не распространяется на произведения научно-популярного характера⁴², в которых вполне могут употребляться языковые средства адресации (в особенности если текст предназначен для детской аудитории), ср.: *если вы помните, как вы знаете, вы можете возразить, видите, по-*

⁴² Как известно, статус научно-популярной литературы остается неопределенным: разными исследователями она понимается то как отдельный стиль, то как подстиль. Обсуждение этого вопроса не входит в рамки данного исследования.

нимаете и пр. [Нистратова 1997]. Помимо этого, указанные местоимения и глагольные формы вполне можно встретить в цитируемых источниках, а также языковых иллюстрациях (в лингвистических текстах). Правда, и те и другие имеют графическое выделение (кавычками или шрифтом), но для того чтобы это отследить, требуется учитывать этот вид информации — наряду с лексической и грамматической, которые традиционно попадают в поле внимания исследователей.

В этой связи мы предлагаем оригинальный подход к распознаванию центральных жанров научной речи (монография, статья, диссертация), основанный на учете графематической информации, приписанной атомам, токенам и их группам на начальном этапе анализа. Наша идея заключается в том, что отличительной особенностью этих жанров является большое разнообразие используемых символов, как буквенных (различных алфавитов), так и небуквенных: достаточно вспомнить оформление ссылок на литературу, сноски, аббревиатуры, цитаты, формулы, маркированные и нумерованные списки, библиографию и пр. Вести учет типов символов возможно благодаря отличительной особенности описываемого лингвистического процессора, связанной с накоплением информации (графематической, морфологической, семантической, синтаксической) по ходу анализа, при движении от одного этапа к другому. Тот факт, что на самом последнем и наиболее сложном (коммуникативно-синтаксическом) этапе оказывается востребованной графематическая информация (начальный этап анализа), представляется нам, помимо прочего, свидетельством правильности выбранной идеологии построения лингвистического процессора⁴³.

Официально-деловой стиль

Официально-деловой стиль речи, обслуживающий сферу административно-правовой общественной деятельности, сближается с научным стилем по ряду параметров — сложному синтаксису, рас-

⁴³ Развивая эту тему, можно отметить роль тире для идентификации художественного текста на русском языке. Как известно, тире в начале абзаца используется для обозначения реплик персонажей в художественном тексте (в английском языке аналогичную роль выполняют кавычки, у которых более высокий индекс неоднозначности). Поэтому если русский литературный текст насыщен диалогами, благодаря тире в такой специфической позиции он будет успешно распознан как таковой.

пространенности безличных и пассивных конструкций, широкому использованию терминов. Если в научном тексте оценка и экспрессия заметно ослаблены по сравнению с художественным и публицистическим стилями, в официально-деловой речи они, можно сказать, сведены к нулю. В проекции на нашу задачу — построение информационной базы по тексту — это означает, что снимается задача выявления мнений и их носителей.

В литературе отмечается, что текстам официально-делового стиля свойственно точное следование «норме» изложения: объективно-нейтральной (в языковом плане) и линейно-последовательной (в плане композиции) подаче содержания [Одинцов 1980: 139–141]. Норма развертывания текста для многих письменных жанров данного стиля является настолько устойчивой, что существуют фиксированные схемы изложения определенного содержания и готовые стандартные формы не только языкового выражения отдельной мысли, но и развертывания целого текста [Татарникова 2006: 94]⁴⁴. Это, несомненно, облегчает автоматическое распознавание подобных текстов⁴⁵, а также обеспечивает однозначность их интерпретации (ср. [Мурзин 1991: 20–21]).

Лексические и грамматические черты жанров официально-делового стиля с разной степенью детализации описаны в ряде работ, ср. [Кожин, Крылова, Одинцов 1982: 104–108; Веселовская 1990; Гуртуева 1991; Кыркунова 1996; Арефьева 1997; Анисимова 2000; Татарникова 2006].

Ядром официально-делового стиля является законодательный подстиль, характеризующийся внутренней однородностью. Для языка законов (указов, постановлений и т. п.) характерны нейтральность изложения, отсутствие экспрессивно-эмоциональных языковых средств, широкое использование абстрактной лексики, юридической терминологии, высокий уровень формализации. Безличный характер законодательной речи исключает использование в ней личных местоимений, личных форм глаголов 1-го и 2-го лица, повелительного и сослагательного наклонений, вопросительных и воскли-

⁴⁴ Именно среди документных жанров мы встречаем наиболее жестко построенный тип текстов — то, что Шварцкопф [1996: 272] называет «текстом-матрицей».

⁴⁵ Не следует также забывать, что все официальные документы непременно снабжены заголовком, определяющим их жанр (ср. *договор, справка, заявление, приказ, акт, протокол* и т. д.).

цательных предложений. Практически исключено использование глагольных форм прошедшего времени.

Важной особенностью законодательного текста, облегчающей его автоматическое распознавание (но существенно затрудняющей конструктивно-синтаксический анализ!), является сложная композиционная структура, подразумевающая многоступенчатое разбиение на части, разделы, главы, статьи, пункты. Таким образом, и в этом случае формальная (графическая) информация оказывается полезной для идентификации типа текста.

Публицистический стиль

Публицистические тексты в силу своего многообразия также имеют мало черт, которые бы позволили автоматически отличать их от произведений иных стилей. Трюизмом стало известное высказывание В.Г. Костомарова [1971] о чередовании экспрессии и стандарта в газетном тексте, однако не вполне ясно, как это наблюдение можно использовать в контексте автоматического анализа. В качестве другой отличительной особенности публицистической речи обычно упоминается обязательность авторской оценки. Однако это утверждение нуждается в уточнении. Во-первых, как указывалось выше, в официально-информативном подстиле (жанрах хроникальной и короткой информационной заметки) она практически сведена к нулю. Во-вторых, проблема автоматического выделения мнений в компьютерной лингвистике еще далека от решения, причем это справедливо даже в отношении первичной задачи, связанной с распознаванием компьютером того факта, что текст несет в себе оценку. Не следует также забывать, что субъективная модальность широко присуща также художественным и научным текстам.

Как замечает О.В. Дунаевская, «в рамках публицистического стиля есть формы, тяготеющие к официально-деловой речи (заметка), есть — к научной (статья), а есть — и к художественной (очерк, эссе, фельетон); итак, мы видим: функционально-стилистическая “сетка” накладывается на отдельно взятый функциональный стиль, воспроизводя свои черты словно в уменьшенном варианте» [Дунаевская 2010: 60–61]. Вероятность обнаружения специфических черт отдельных жанров выше, чем стиля в целом, хотя и не столь высока, как обнаружилось применительно к отдельным жанрам научной и официально-деловой речи. Это наглядно видно на примере описания [Коньков 2004], посвященного информационным и аналитическим

жанрам газетной речи. По-видимому, сохраняет справедливость и суждение А. Н. Васильевой [1982: 29] о значительном расхождении между жанровым и лингвостилистическим членением.

Примером исследования отдельных жанров публицистической речи может служить статья [Карпилович 2013], где автор обращается к рассмотрению структурных и языковых особенностей проблемной статьи (а также проблемного очерка и интервью). Т. П. Карпилович выделяет присущую данному жанру суперструктуру (в терминах Ван Дейка [1989]), а также ряд лексических маркеров и характерных синтаксических конструкций. Так, она констатирует, что в надзаголовке, заголовке, подзаголовке или в начальных абзацах текста часто присутствуют слова *проблема, вопрос, препятствие, трудность*; при изложении сути проблемы нередко используются отрицательные предикаты, слова с отрицательным значением, которые описывают различного рода запреты, невыполнение каких-либо действий, приводящие к конфликту. Впрочем, автор отмечает необязательность указанных лексических средств. По ее мнению, больший оптимизм с точки зрения возможностей формализации вселяют средства языкового воплощения другого прототипического компонента проблемных очерков и интервью, а именно «пути (способы) решения проблемы». Его индикаторами в первую очередь служат структуры с прямой или косвенной речью, а также характерные лексические комплексы, включающие слова со значением специальностей и должностей, к примеру: *специалист фирмы, начальник департамента по социальному развитию* и т. п. Именно в этой структурной части статьи широко представлена эмотивная лексика, императивные конструкции, в том числе с отрицательными предикатами, риторические вопросы, наречия со значениями важности, необходимости и прочие средства выражения оценки.

Эти наблюдения представляют интерес, однако возможность их использования для автоматической идентификации произведений рассматриваемого жанра вызывает сомнения (сама автор статьи преследует иную цель, а именно автоматическое реферирование). Прежде всего, они связаны с тем, что, «в отличие от композиции, компоненты суперструктуры часто не присутствуют в тексте в виде отдельных законченных фрагментов, а находятся в различных частях текстового пространства» [Карпилович 2013: 6]. Это фактически сводит к нулю ценность выделения структурных компонентов и характерных для них языковых средств. Таким образом, в «сухом остатке» мы имеем некоторый необязательный перечень типов

лексических единиц и синтаксических конструкций, и трудно сказать, насколько надежно они могут служить индикаторами рассматриваемого жанра. Безусловно, это объясняется спецификой публицистического жанра в целом и никоим образом не может считаться недостатком обсуждаемого исследования.

* * *

Наиболее сложную задачу представляет собой идентификация художественного текста. По-видимому, в отношении него в общем случае⁴⁶ следует применять принцип «от противного», а именно: гипотеза о принадлежности текста к этому стилю выдвигается тогда, когда проверка по всем чертам, характерным для текстов других типов, дает отрицательный результат (т.е. они либо не представлены вовсе, либо соответствующие подсчеты не превышают средних значений).

***Функционально-смысловые типы речи:
описание, повествование, рассуждение***

В зависимости от темы сообщения и коммуникативных намерений речевая деятельность организуется так, чтобы логико-понятийные свойства сообщаемой информации соответствовали намерениям говорящего.

Тема сообщения, являясь существенным и необходимым признаком всякого связного текста, составляет ядро содержания текста и определяет его структуру. Тематику конкретного текста можно определить по объекту высказывания, ответив на вопрос «о чем здесь говорится?».

Коммуникативные намерения говорящего различаются по способам изложения информации, называемыми композиционно-речевыми формами или функционально-смысловыми типами речи [Нечаева 1974, 1999; Брандес 1990; Всеволодова 2000; Валгина 2003: 75–96] (Стилистический энциклопедический словарь... 2006). В тексте их можно определить, ответив на вопросы «как излагается информация?», «о чем говорится?», «какие информационные задачи решает автор?». В зависимости от ответов на эти вопросы выделяется три функционально-смысловых типа: описание, повествование и рассуждение.

⁴⁶ За исключением текстов, насыщенных диалогами (см. выше), а также драматургических произведений.

Каждый из названных типов речи⁴⁷ имеет отличительные черты по речевым свойствам используемого арсенала языковых средств (функционально-семантические признаки) и применяемым логикориторическим способам и приемам изложения информации (понятийно-смысловые признаки).

С логической точки зрения типы речи дифференцируются следующим образом: для сообщения характерны диахронические отношения между признаками объекта (развивающиеся, изменяющиеся действия или состояния), для описания — синхронные отношения (одновременные признаки объекта), для рассуждения — каузальные (причинно-следственные отношения между суждениями об объекте).

В функциональном и смысловом отношении описание сводится к ответам на вопрос «какой объект обсуждается?», «какие стороны объекта обсуждаются?». Тип описание означает, что в тексте содержится фактическая констатирующая характеристика сосуществующих объектов, свойств предметов, наличных явлений и ситуаций. Описание используется для статического изображения действительности, для констатации состояния действительности, местности, интерьера, важных черт личности, характера или внешности человека.

Типологическим подтипом описания является разъяснение. Разъяснение используется говорящим для истолкования содержания какого-либо текста, объяснения совокупного смысла какого-либо события, какой-либо ситуации на основе личного опыта и обусловленного знаниями истолкователя.

Как в описании, так и в разъяснении используются по преимуществу логические и аналитические средства языка, обращенные к разуму (к рассудку). Композиционная схема описания выглядит следующим образом:

- 1) указание на объект описания (тема описания);
- 2) выделяемые денотативные признаки объекта описания;
- 3) характеристика свойств, оценка качеств объекта описания.

⁴⁷ Помимо этих трех способов изложения, известных еще из старинных риторик, современные авторы выделяют и другие, дополнительные — объяснение [Вейзе 1985], доказательство [Цветкова 1982]. Еще больше различных типов речи предлагается в работе [Одинцов 1980]; расширение основного списка поддерживает и Н. С. Валгина, указывая на то, что обозначенных выше трех типов оказывается явно недостаточно, когда мы выходим за рамки художественных произведений [Валгина 2003: 92–93]. Так, в диалоге нередко вопросы и побуждения к действию. Наш подход выводит такие случаи за рамки понятия функционально-смысловых типов речи, и мы считаем возможным ограничиться классическими тремя типами.

Повествование, отвечая на вопросы типа: «что происходит или произошло с объектом повествования?», «какие события повлияли на объект повествования?», отражает действительность в динамике, характеризует объекты, события во временной последовательности и развитии. Коммуникативные намерения, реализующиеся в описаниях и повествовании, обычно облакаются в форму дескриптивных высказываний. Композиционная схема повествования следующая:

- 1) начало события, в которое вовлечен объект информации, завязка;
- 2) развитие события;
- 3) кульминация события;
- 4) конец события, развязка.

В основе рассуждения лежат ответы на вопросы: «почему выбран данный объект рассуждения?», «почему данный объект таков?», «почему нечто произошло/происходит с объектом рассуждения?», «каковы причины изучаемого объекта, явления (события, действия)?», «каковы следствия из данного явления?», «что значит для нас изучаемый объект, изучаемое явление (событие, действие)?».

Этот функционально-смысловой тип речи используется при описании причинно-следственных связей явлений и событий действительности, с его помощью производится обобщение фактов, устанавливаются имплицативные отношения между событиями и их рациональной оценкой. Рассуждения составляют базу нормативных высказываний, т. е. устанавливающих правила и ценности. В рассуждении как функционально-смысловом типе речи используются по преимуществу логические и аналитические средства убеждения, обращенные к разуму (к логосу).

Рассуждение строится по следующей композиционной схеме:

- 1) тезис, т. е. суждение, обосновываемое в процессе аргументации;
- 2) доказательства, аргументы, подтверждаемые примерами;
- 3) вывод или заключение (может отсутствовать в тексте).

Очевидно, что многие речевые жанры допускают чередование разных типов речи⁴⁸. Так, в научных текстах часто встречается опи-

⁴⁸ Считаем нужным подчеркнуть, что здесь и далее выражение «типы речи» следует понимать именно в вышеуказанном смысле, поскольку в литературе встречаются и другие его трактовки, ср. [Шмелев 1977; Кожин, Крылова, Одинцов 1982; Солганик 2010: 80–94].

сание объекта и рассуждение, но может присутствовать и повествование (например, если речь идет о процедуре экспериментального исследования). В публицистике, за редким исключением, обязательным компонентом является рассуждение, но оно, в свою очередь, базируется на некоем описании объекта или сообщении о событии. В художественном произведении чаще всего перемежаются сообщение и описание, но встречается и рассуждение.

Рассмотрим, что представляют собой данные типы речи с точки зрения построения информационной структуры текста. Начнем с описания (дескриптива). Согласно О. А. Нечаевой, описание — это «модель монологического сообщения в виде перечисления одновременных или постоянных признаков предмета в широком понимании» [Нечаева 1974: 94]. В таких текстах референтная функция отчетливо доминирует над остальными [Чебанов, Мартыненко 1999: 6]. В литературе выделяются разные виды описания в зависимости от типа ремы. Наиболее распространенными являются описания с качественной и предметной ремой [Валгина 2003: 80–85].

Описания с качественной ремой статичны: внимание автора фиксируется на свойствах одного предмета, который является темой. Такого рода описания нередко встречаются в научно-популярных изданиях, справочниках, энциклопедиях, ср.:

Марс — четвертая от Солнца планета Солнечной системы. Она хорошо заметна невооруженным глазом как весьма яркая звезда, имеющая максимальную звездную величину $-2,0^m$. В своем видимом движении Марс довольно быстро перемещается по небу, поскольку его орбита расположена близко к Земной.

Диаметр Марса вдвое меньше земного, а его масса в десять раз меньше массы нашей планеты. Тонкая марсианская атмосфера сильно разрежена. Это позволяет наблюдать поверхность планеты с помощью телескопов. На ней хорошо различимы темные и светлые участки, горы, ущелья, а также знаменитые полярные шапки⁴⁹.

Другой характерный пример описания с качественной ремой заимствован нами из путеводителя:

Родос, главный город острова, существует на самой его северной окраине с 408 г. до н.э. С трех сторон город омывается морем. Вдали почти всегда ясно видно побережье Малой Азии. Длинный галечный пляж, часто обдуваемый сильными ветрами, окаймляет западное по-

⁴⁹ Планеты и созвездия: иллюстрированный справочник. Вильнюс, 2013. С. 94.

*бережье. На севере почти вплоть до городского центра простирается песчаный пляж*⁵⁰.

Описаниям с качественной ремой соответствуют следующие построения в информационной структуре. Создается объект — тема текста (в наших примерах — *Марс, Родос*), и ему последовательно, по мере обработки предложений, приписываются признаки. (Для нас пока достаточно весьма общего, грубого представления, не учитывающего различных нюансов построения, которые могут быть связаны, например, с тем, что разные признаки могут характеризовать объект не целиком, а какие-то его отдельные части и аспекты.)

Описания с предметной ремой также предполагают статичное положение наблюдателя, который как будто смотрит вокруг себя и приписывает характеристики отдельным объектам, составляющим «компоненты» того места, в котором он находится, ср.:

В то лето я жил в маленьком северном городе. Город стоял на берегу реки. По реке плыли белые пароходы, грязно-бурые баржи, длинные плоты, ширококулые карбасы, с запачканными черной смолой бортами. У берега стояла пристань, пахнувшая рогожей, канатом, сырой гнилью и воблой. На пристани этой редко кто сходил, разве только пригородные колхозники в базарный день да унылые командировочные в серых плащах, приезжавшие из области на лесозавод.

Вокруг города по низким пологим холмам раскинулись леса, могучие, нетронутые: лес для сплава рубили в верховьях реки. В лесах попадались большие луговины и глухие озера с огромными старыми соснами по берегам. Сосны все время тихонько шумели. Когда же с Ледовитого океана задувал прохладный влажный ветер, нагоняя тучи, сосны грозно гудели и роняли шишки, которые стукались о землю (Ю. Казаков. Арктур — гончий пес).

Такие описания предполагают иные построения в информационной структуре, чем те, что были указаны выше для описаний с качественной ремой. Здесь теме соответствует некий «суперобъект» — в данном случае *город*. Ему подчинены прочие объекты — *пароходы, пристань, леса, озера, сосны* и пр., каждому из которых приписывается свой (-и) признак (-и)⁵¹. Встает задача моделирования

⁵⁰ Родос: путеводитель (серия «Nelles Verlag»). СПб., 2010. С. 17.

⁵¹ Проблема выделения подобных «суперобъектов» в тексте и построения соответствующих иерархических связей в информационной структуре далека от

пространственных связей между всеми этими объектами. Требуется также включить в эту картину фигуру повествователя, которая в нашей информационной структуре наделена особым статусом.

Независимо от типа ремы все создаваемые признаки имеют статус 'одновременный'. Что касается того, следует ли считать эти признаки постоянными, заслуживает внимания наблюдение В. А. Яцко о различиях между описанием в научной и художественной речи. По его словам, в отличие от описания в научном тексте, описание в художественном произведении характеризуется тем, что «наряду с постоянными, повторяющимися признаками объекта зачастую указываются и временные признаки, непосредственно наблюдаемые в момент речи...» [Яцко 1998: 129–130].

Важная особенность описания как типа речи — нейтральная модальность, невозможность выражения модуса мнения [Яцко 1998: 127]. Следовательно, здесь не встает задача распознавания мнений (*opinion mining*) и привязки пропозиций к носителям соответствующих точек зрения. Правда, сказанное безоговорочно верно лишь для научного описания: в публицистике и художественном тексте сам выбор слов, осуществляющий категоризацию действительности, несет эмоциональную нагрузку, так что здесь можно говорить о другой, более тонкой и сложной проблеме — определении тональности текста (*sentiment analysis*).

Рассматривая языковые индикаторы описания как типа речи, сошлемся на наблюдение О. А. Нечаевой, которая пишет: «Описанию не свойственна косвенная, ирреальная модальность (со значением пожелания, условия) в форме сослагательного наклонения или значение побуждения, просьбы, т. е. повелительное наклонение, наконец, невозможна и предписывающая модальность в виде необходимости, долженствования» [Нечаева 1974: 44]. Весьма редко используются в описании и вводные слова [Нечаева 1974: 45]. В целом «ограниченное использование в описании вводных слов, как и наклонений гла-

решения, и дело не упирается здесь исключительно в разработку энциклопедического компонента, содержащего знания о мире. Дело в том, что многие отношения, которые в тексте оказываются иерархическими, не являются таковыми в действительности: так, *пароходы, плоты, вобла, колхозники* не составляют непрерывный атрибут городского пейзажа. В принципе, никакой энциклопедический компонент не способен вместить в себя все то, что может быть упомянуто в тексте, когда речь идет о городе. Сказанное, разумеется, справедливо и в отношении других объектов: посредством ассоциативных связей практически все может быть связано со всем.

гола, по-видимому, объясняется одной причиной: констатирующей сущностью описательной речи» [Нечаева 1974: 46]. Вместе с тем, как видно из приведенных формулировок, речь идет не о жестких правилах, а закономерностях, и сама автор на страницах своей книги приводит примеры, идущие вразрез с указанными тенденциями.

Из характерных синтаксических особенностей описания как типа речи следует отметить номинативные, безличные и эллиптические предложения, характеризующие место, время, состояние окружающей среды и пр. [Нечаева 1974: 82–90]. Впрочем, эти филологические наблюдения, к сожалению, не подтверждены статистически.

Заслуживает внимания упоминание о том, что художественное описание чаще строится в прошедшем времени, а научное или деловое — в настоящем [Нечаева 1974: 50]. Впрочем, это справедливо скорее в отношении двух последних стилей. В художественном же описании нередко сочетаются формы настоящего и прошедшего, глаголы несовершенного и совершенного вида. Иными словами, такие формальные показатели, как смена глагольного вида или времени, не являются надежным индикатором смены типа речи (см. также ниже). В качестве иллюстрации используем пример из цитируемой монографии [Нечаева 1974: 52]:

*Бричка бежит, а Егорушка **видит** все одно и то же — небо, равнину, холмы... Музыка в траве **притихла**, старички **улетели**, куропаток не видно. Над поблекшей травой... **носятся** грачи: все они похожи друг на друга...* (А. П. Чехов. Степь).

Пространные описания демонстрируют весьма прихотливое переплетение временных пластов и, соответственно, личных форм глагола. Это справедливо и в отношении тех художественных произведений, которые целиком построены в соответствии с данным типом речи. Таков, к примеру, рассказ И. А. Бунина «Антоновские яблоки». Вот один лишь абзац:

*Из такой трепки сад **выходил** почти совсем обнаженным, засыпанным мокрыми листьями и каким-то притихшим, смирившимся. Но зато как красив он **был**, когда снова **наступала** ясная погода, прозрачные и холодные дни начала октября, прощальный праздник осени! Сохранившаяся листва теперь **будет висеть** на деревьях уже до первых заморозков. Черный сад **будет сквозить** на холодном бирюзовом небе и покорно ждать зимы, пригреваясь в солнечном блеске. А поля*

уже резко **чернеют** пашнями и ярко **зеленеют** закустившимися озимями... Пора на охоту! (И. А. Бунин. Антоновские яблоки).

* * *

Сообщение (повествование, нарратив) рассказывает о некоторой последовательности действий или состояний, поэтому объектам в информационной структуре должен автоматически приписываться статус 'временный'. Далее встает вопрос об отношении данного признака к предшествующим и последующим признакам того же объекта. Отменяет он предыдущий (*поплакал и успокоился*) или нет (*лег и уснул*)? Отменяется ли он последующим? Строго говоря, поскольку анализ осуществляется поступательно, достаточно всякий раз задаваться лишь первым вопросом, т.е. «оглядываться назад», соотносить данный признак с предшествующим, уже имеющимся в базе (если он есть).

Сам по себе это вопрос непростой, но нужно понимать, что основная проблема лежит гораздо глубже. Ситуация, в принципе, напоминает ту, с которой мы столкнулись ранее, рассматривая темпоральную организацию различных речевых жанров (см. параграф «Определение временной структуры текста»). Мы исходили из того, что жанр нам известен, и лишь позднее признались, что распознавание жанра — самостоятельная сложная задача. Так и здесь: мы начали говорить о временных границах признаков в нарративе, но откуда лингвистический процессор знает, что имеет дело с нарративом? Единственное, что доступно наблюдению, — это языковые формы, от них и следует отталкиваться. Перестроим наши рассуждения в этом ключе.

Фундаментальная проблема заключается в том, что нарратив в русском языке может задействовать разнообразные грамматические формы глаголов (как совершенного, так и несовершенного вида), причастий, деепричастий, а также существительные признаковой семантики — отсюда трудность распознавания этого типа речи. Рассмотрим возможные комбинации, отвлекаясь от лексического значения словоформ и принимая во внимание лишь грамматический вид.

Если личному глаголу совершенного вида в прошедшем времени предшествует другой такой глагол (обозначим эту ситуацию как СВ + СВ), то в тексте мы имеем нарратив, характеризующийся иконическим отражением событий. Семантика этих форм позволяет создать в повествовательном тексте цепь событий, характеризую-

щихся, согласно В. Шмиду [2003], непредсказуемостью, необратимостью, неповторяемостью и консекутивностью. При построении информационной структуры происходит поочередное создание соответствующих признаков. При этом каждый предыдущий сохраняет свою значимость свершившегося факта (ср. *Пришел, увидел, победил*). По-видимому, это верно и для деепричастий совершенного вида и страдательных причастий прошедшего времени.

Вариант НСВ + СВ не столь однозначен. Довольно часто здесь, как и в предыдущем случае, появление глагола совершенного вида в прошедшем времени влечет вытеснение предшествующего признака, как, скажем, в конструкции *говорил долго и внезапно умолк*. Однако это не всегда так, ср. *ехал в автобусе и вдруг подумал, лежал в гамаке и заметил*.

Когда второй признак выражен личным глаголом несовершенного вида прошедшего времени, невозможно однозначно утверждать, имеем ли мы дело с синхронными или сменяющимися друг друга признаками. Можно утверждать, что вариант СВ + НСВ часто связан со сменой типа речи, а именно переходом от сообщения к описанию (примеры см. ниже). Что касается последовательности НСВ + НСВ, она встречается как в дескриптиве, так и в нарративе (*подхожу и спрашиваю*), выражать как синхронные, так и последовательные признаки. Это связано с семантикой форм прошедшего несовершенного, которые могут служить как для описания отдельной ситуации, «наличной к определенному моменту в течении событий» [Маслов 1984: 183], так и для описания регулярно повторявшихся в прошлом ситуаций (в том числе событийного плана). Следовательно, в этих случаях первому глаголу не может быть приписан какой бы то ни было определенный статус.

В теоретическом плане граница между дескриптивом и нарративом не является четкой из-за промежуточного положения, которое занимают, во-первых, динамические описания и, во-вторых, рассказы о рутинных событиях, лишенные осложнения (то, что в англоязычной традиции именуется *habitual narratives*). Однако для нашей задачи однозначное отнесение к той или иной категории несущественно; важно то, какие построения в информационной структуре влечет тот или иной тип текста. Обратимся к примерам.

Спорный статус динамического описания обусловлен тем, что в нем используются глаголы активного действия, но они служат не для рассказа о реальных действиях, совершаемых во времени, а для

характеристики лица, предмета и др. Динамическое описание, как правило, имеет предметную рему, ср.:

*Окно моей комнаты выходило в одичавший сад, заросший смородиной, малиной, лопухом и крапивой вдоль забора. По утрам за окном **возились** воробьи, тучами **налетали** дрозды клевать смородину, доктор не **гонял** их и ягоду не **собирал**. На забор иногда **взлетали** соседские куры с петухом. Петух громогласно **пел**, **вытягивая** кверху шею, **дрожал** хвостом и с любопытством **смотрел** в сад. Наконец он не выдерживал, **слетал** вниз, за ним **слетали** куры и поспешно **начинали рыться** возле смородиновых кустов. Еще в сад **забредали** коты и, затаясь возле лопухов, **следили** за воробьями (Ю. Казаков. Арктур — гончий пес).*

Собственно говоря, с точки зрения построения информационной структуры по тексту разграничение между статическим и динамическим описанием несущественно: и то, и другое предполагает приписывание признаков объекту. Более значимым является рассмотренное выше различие между качественной и предметной ремой, связанное с числом создаваемых объектов. Если рема предметная, требуется создавать более одного объекта и, как указывалось выше, встает задача выявления имплицитных связей между ними в пределах одного «суперобъекта».

Нарративы без осложнения, повествующие о регулярно повторяющихся ситуациях, широко распространены в художественной литературе. Поскольку они основываются на использовании глаголов несовершенного вида, которые могут служить и для описания отдельной ситуации, возникает непростая задача автоматической дифференциации нарратива без осложнения и «обычного» повествования.

В отличие от описания (дескриптива), в повествовании (нарративе) нередко вводные слова со значением оценки, а также обстоятельственные слова со значением временной последовательности, призванные отражать динамику повествования [Нечаева 1974: 128–132].

* * *

Отличительная черта рассуждения как типа речи — насыщенность текста субъективной модальностью, что особенно характерно для публицистического стиля. Именно здесь наиболее остро стоит уже неоднократно упоминавшаяся проблема разграничения факта

и мнения. Создаваемые в информационной структуре пропозиции, отражающие диктум (по Ш. Балли), должны снабжаться привязкой к соответствующему модусу и его автору. Это специальная задача, которая будет рассмотрена ниже.

Если для публицистического стиля основная оппозиция типов речи — это рассуждение vs. (описание + сообщение), то для художественного важнее всего разграничивать описание и сообщение. Именно на этой проблеме мы хотим остановиться подробнее. Обратимся к распространенной ситуации, при которой дескриптивные фрагменты (описания персонажей, предметов обстановки, места действия и пр.) вкраплены в нарративный художественный текст. Возникает вопрос: как можно автоматически распознать, где в литературном произведении кончается описание и начинается действие (или наоборот)? Как уже указывалось, от ответа на этот вопрос зависит статус создаваемых в информационной структуре признаков: одновременный (синхронный) или последовательный. Проанализируем, какие языковые средства служат маркерами перехода от нарратива к дескриптиву и обратно⁵².

Пожалуй, самым частотным показателем является смена глагольного вида (если и повествование, и описание строятся в прошедшем времени). В приводимом ниже отрывке можно видеть переход от описания обстановки (глаголов несовершенного вида) к действию (глаголам совершенного вида) и затем к описанию персонажа (снова глаголам несовершенного вида), ср.:

*В горнице **было** тепло, сухо и опрятно: новый золотистый образ в левом углу, под ним покрытый чистой суровой скатертью стол, за столом чисто вымытые лавки; кухонная печь, занимавшая дальний правый угол, ново **белела** мелом; ближе **стояло** нечто вроде тахты, покрытой пегими попонами, упиравшейся отвалом в бок печи; из-за печной заслонки сладко **пахло** щами — разварившейся капустой, говядиной и лавровым листом.*

*Приезжий **сбросил** на лавку шинель и **оказался** еще стройнее в одном мундире и в сапогах, потом **снял** перчатки и картуз и с усталым видом **провел** бледной худой рукой по голове — седые волосы его с чечесами на висках к углам глаз слегка **курчавились**, красивое удлиненное*

⁵² Заметим, что графический признак (начало нового абзаца) ненадежен: в общем случае он не является ни необходимым, ни достаточным условием смены типа речи.

лицо с темными глазами **хранило** кое-где мелкие следы оспы (И. А. Бунин. Темные аллеи).

Переход от описания к действию могут сигнализировать такие лексические средства, как *однажды, случай, случилось, неожиданно* и пр. Так, длинный абзац, посвященный наружности и характеру Сильвио в начале пушкинской повести «Выстрел», заканчивается фразой *Нечаянный **случай** всех нас изумил*, — а следующий абзац, в котором речь идет о происшествии за карточным столом, открывается предложением ***Однажды** человек десять наших офицеров обедали у Сильвио*. В «Капитанской дочке» окончание описания детства главного героя обозначается фразами *Тут **судьба** моя **переменилась***. ***Однажды**...* Обратный переход от нарратива к описанию не имеет регулярных лексических средств выражения и обычно маркируется только сменой глагольного вида.

Кроме смены глагольного вида, в художественной литературе также широко представлено переключение с прошедшего времени повествования на настоящее время описания (и обратно). В частности, этот прием регулярно используется для характеристики географической местности:

*Соден — небольшой городок в получасовом расстоянии от Франкфурта. Он **лежит** в красивой местности, на отрогах Таунуса, и известен у нас в России своими водами, будто бы полезными для людей с слабой грудью. Франкфуртцы **ездят** туда большие для развлечения, так как Соден **обладает** прекрасным парком и разными «виртишафтами», где можно пить пиво и кофе в тени высоких лип и кленов. Дорога от Франкфурта до Содена **идет** по правому берегу Майна и вся обсажена фруктовыми деревьями.*

*Пока карета тихонько катилась по отличному шоссе, Санин украдкой **наблюдал** за тем, как Джемма **обращалась** со своим женихом: он в первый раз **видел** их обоих вместе (И. С. Тургенев. Вешние воды).*

При описании персонажа подобное переключение тоже встречается, но значительно реже. В связи с тем, что интересующий нас фрагмент достаточно длинный, приводим его с сокращениями:

*Оборачиваюсь: Грушницкий! Мы **обнялись**. Я **познакомился** с ним в действующем отряде. Он **был ранен** пулей в ногу и **поехал** на воды с неделю прежде меня.*

*Грушницкий — юнкер. Он только год в службе, **носит**, по особенному роду франтовства, толстую солдатскую шинель. У него георгиев-*

ский солдатский крестик. Он хорошо сложен, смугл и черноволос; ему на вид можно дать двадцать пять лет, хотя ему едва ли двадцать один год. Он **закидывает** голову назад, когда говорит, и поминутно **крутит** усы левой рукой, ибо правой **опирается** на костыль. **Говорит** он скоро и вычурно <...>

Впрочем, в те минуты, когда **сбрасывает** трагическую мантию, Грушницкий довольно мил и забавен. Мне любопытно видеть его с женщинами: тут-то он, я думаю, **стареется!**

Мы **встретились** старыми приятелями. Я **начал** его расспрашивать об образе жизни на водах и о примечательных лицах (М. Ю. Лермонтов. Герой нашего времени).

Заметим, что схожий механизм может использоваться также при переходе от нарратива к рассуждению и обратно, ср.:

Он на пороге **остановился**: ему **хотелось** пожать мне руку... И если б я показал ему малейшее на это желание, то он бросился бы мне на шею; но я **остался** холоден, как камень — и он **вышел**.

Вот люди! все они таковы: **знают** заранее все дурные стороны поступка, **помогают**, **советуют**, даже **одобряют** его, видя невозможность другого средства, — а потом **умывают** руки и **отворачиваются** с негодованием от того, кто имел смелость взять на себя всю тяжесть ответственности. Все они таковы, даже самые добрые, самые умные!..

На другой день утром, получив приказание от высшего начальства отправиться в крепость N, я **зашел** к княгине проститься (М. Ю. Лермонтов. Герой нашего времени).

Однако в целом этот признак ненадежный, так как в пределах одного нарративного фрагмента возможно сочетание форм прошедшего времени с настоящим или будущим в стилистических целях — для оживления излагаемых событий, для подчеркивания длительности действия и пр.

Бывает, что переходы от нарратива к дескриптиву и назад не маркируются ни сменой глагольного вида, ни времени. Что же в таком случае выступает носителем соответствующей информации? Как читающий текст человек понимает, что такой переход имеет место? Есть примеры, в которых соответствующим сигналом является употребление глагола с иными семантическими характеристиками, чем предшествующие глаголы. Так, в примере ниже глаголы движения *идти* и *кланяться* употребляются в нарративном фрагменте, а статический глагол *быть* — при описании внешности персонажа, ср.:

Она **шла** босая, с дерюжной сумой через плечо, подпираясь высокой палкой, и в деревьях и селах молча **кланялась** перед каждой избой. Мальчик **шел** за ней сзади, тоже с мешком через плечико, в старых башимаках ее, разбитых и затвердевших, как те опорки, что валяются где-нибудь в овраге.

Он **был** урод. У него **было** большое, плоское темя в кабаньей красной шерстке, носик расплюснутый, с широкими ноздрями, глазки ореховые и очень блестящие. Но когда он улыбался, он **был** очень мил (И. А. Бунин. Дурочка).

Однако это наблюдение не следует возводить в общее правило. Выше уже упоминалось о том, что в художественном тексте одни и те же глаголы могут использоваться как в нарративе, так и в дескриптиве⁵³. Показателен, например, следующий фрагмент, где позиция с глаголом *вливался* завершает нарратив о продвижении судна, а последующий текст посвящен описанию, но также преимущественно посредством глаголов действия. Ср.:

Пароход медленно **вливался** в Азию — в приторный дым кизяка, в туманы пыли, в огни прибрежный костров, в кочевья, раскинутые вокруг города, в заунывные **крики** киргизов, **тянувших** на песке мокрые невода.

Шелудивые псы **бежали** по берегам с хриплым лаем, за ними **мчались** голые киргизские дети. Пестрое тряпье, развешанное на шестах, **приветствовало** нас подобно знаменам Тамерлановых полчищ. **Рыдали**, как двери на тугих пружинах, облезлые верблюды.

По берегам **брели** киргизы, **волоча** хвостами по пыли **трепещущих** сазанов и судаков. По их грязной чешуе стекали чистые капли крови.

Чигири **скрежетали** во внезапной темноте, похожей на кофейную гущу. Огненная надпись «Эмбанефть», висящая в воздухе, как бы **открывала** вход в тяжелую пустыню (К. Г. Паустовский. Подводные ветры).

Разумеется, можно трактовать три последних абзаца как нарратив (или скорее цепь из нескольких нарративов: про псов, верблюдов, киргизов и пр.), но тогда встает вопрос о соотносительности между

⁵³ Ср.: «Факты свидетельствуют о том, что семантика глаголов, указывающих на состояние предмета или его действие, движение, не соответствует понятиям статичного в описании и динамичного в повествовании. Глагол со значением действия, даже движения, может передавать статичность содержания в описании, а глагол состояния — динамику повествования» [Нечаева 1974: 114].

собой этих нарративов, в том числе отношениях вложенности. Заметим, что приведенный отрывок не является каким-то редким, исключительным или присущим авторскому стилю конкретного писателя. Вновь обратившись к творчеству Бунина, можно найти схожий пример, в котором глаголы действия также используются для описания обстановки. Переход к нарративу маркируется здесь словом *вдруг* (аналогичную функцию могли бы также выполнять обороты *и тут, в это мгновение* и т. п.):

*В глаза **бьет** зарево пожара вдали направо. Тележка стоит против того леска, что открывался при зарницах. Лесок от зарева стал теперь черным и весь зыбко **дрожит**, как **дрожит** и все поле перед ним в сумрачно-красном трепете от того жадно несущегося в небе пламени, которое, несмотря на даль, **попыхает** с бегущими в нем тенями дыма точно в версте от тележки, **разъяряется** все жарче и грознее, **охватывает** горизонт все выше и шире, — кажется, что жар его уже **доходит** до лица, до рук, виден даже над чернотой земли красный переплет какой-то сгоревшей крыши. А под стеной леса стоят, багрово серея, три больших волка, и в глазах у них **мелькает** то сквозной зеленый блеск, то красный, — прозрачный и яркий, как горячий сироп варенья из красной смородины. И лошади, шумно всхрапнув, **вдруг** диким галопом **ударяют** вбок, влево, по пашне, малый, на вожжах, **валится** назад, а тележка, со стуком и треском, мотаясь, **бьется** по взметам... (И. А. Бунин. Волки).*

Напоследок заметим, что в художественной литературе повествование и описание иногда столь тесно переплетены, что граница между ними идет по отдельным предложениям внутри абзаца (если не по частям одного предложения). В таких случаях говорят о «контаминации функционально-смысловых типов речи» [Нечаева 1974: 195–203], или о «сложных смешанных типах речи» [Валгина 2003: 94], ср.:

Я пошел вправо через кусты [повествование]. Между тем ночь приближалась и росла, как грозовая туча; казалось, вместе с вечерними парами отовсюду поднималась и даже с вышины лилась темнота [описание]. Мне попала какая-то неторная, заросшая дорожка; я отправился по ней, внимательно поглядывая вперед [повествование]. Все кругом чернело и утихало, одни перепела изредка кричали [описание]. Небольшая ночная птица, неслышно и низко мчавшаяся на своих мягких крыльях, почти наткнулась на меня и пугливо нырнула в сторону. Я вышел на опушку кустов и побрел по полю межой

[повествование]. Уже с трудом различал я отдаленные предметы; поле неясно белело вокруг; за ним, с каждым мгновением надвигаясь громадными клубами, вздымался угрюмый мрак. Глухо отдавались мои шаги в застывающем воздухе. Побледневшее небо стало опять синеть — но то уже была синева ночи. Звездочки замелькали, зашевелились на нем [описание] (И. С. Тургенев. Бежин луг).

* * *

Проблема разграничения повествования и описания встает не только при анализе художественной литературы. Выше мы рассматривали ситуацию, где в нарративный текст вставлены дескриптивные фрагменты. Для полноты картины обратимся теперь к противоположному случаю: в дескриптивный текст включен нарративный фрагмент. Это широко встречается, например, в путеводителях. Нарратив здесь либо служит для рассказа об истории возникновения достопримечательности (в этом случае наблюдается переключение временных форм глагола), либо обозначает типовой маршрут путешественника. Приведем примеры обоих случаев.

Костел Св. Микулаша **создан** двумя выдающимися пражскими зодчими — Кристофом и Килианом Игнацом Динценхоферами. Отец, Кристоф, **построил** в начале XVIII столетия неф и западный фасад, а сын спустя 40 лет хор и купол. **Завершил** постройку созданием колокольни (1755) Ансельмо Лураго. Размеры костела **впечатляют**: его куполу на высоте 75 метров могла бы позавидовать смотровая площадка на холме Петршин. Не **имеют** себе равных и росписи сводов главного нефа «Апофеоз Св. Микулаша» (Я. Л. Кракер), занимающие площадь 1500 м². В интерьере **обращают** на себя внимание живопись К. Шкреты и скульптура И. Ф. Платцера⁵⁴.

Если нарративный фрагмент посвящен изложению маршрута путешествия, он может быть оформлен в настоящем времени, как и дескриптивный текст. Переход от статического описания достопримечательности к (воображаемому) движению можно распознать по широкому использованию лексики движения, а также частым отсылкам к эгоцентрической системе пространственных координат, ср.:

Пропилеи — парадный вход на Акрополь, а само слово можно примерно перевести как «преддверие». Второе по времени постройки здание

⁵⁴ Прага: путеводитель (серия Polyglott). СПб., 1996. С. 56.

на Акрополе после Парфенона сооружено по проекту Мнесикла в 437–432 годах до н. э. Задача здания — стать архитектурным предисловием к Парфенону, мощью и гармонией создавать торжественное настроение поднимающемуся на Акрополь. С этой задачей постройка Мнесикла, несмотря на разрушения, великолепно справляется до сих пор.

Посетитель **подходит** к Пропилям сбоку и внезапно **оказывается** на широкой мраморной лестнице, охваченной двумя крыльями-портиками. По мере **приближения** колоннады крыльев **раступаются** и уступают все поле зрения центральному фасаду из дорических колонн, которые, в свою очередь, **раскрываются** посередине и уводят внутрь здания, в коридор, обрамленный ионическими колоннами. Коридор **приводит** к сверхчеловеческих размеров центральной двери, сразу за которой — внутренний фасад из дорических колонн. **Пройдя** под колоннами, **оказываетесь** на Священной дороге с парадным видом на Парфенон. Рядом с центральной «дверью для богов» — еще четыре, уменьшающиеся от центра к краям, для людей.

Левое крыло Пропилей занимала Пинаотека — картинная галерея. Это одно из первых известных нам собраний живописи⁵⁵.

Изложение маршрута может маркироваться и другими средствами, в числе которых следует упомянуть переключение с нейтрального, «безличного», рассказа на персонализированное мы-или вы-повествование, а также использование директивных речевых актов, ср.:

Отдохнув, **спустимся** к Висле, **полюбимся** ее берегами с обзорной площадки, а дальше **направимся** к городским воротам — барбакану⁵⁶.

В Новом городе **вы повстречаете** огромное количество костелов: костел Святого Духа, костел Святого Казимира и многие другие⁵⁷.

Зайдите в новое здание библиотеки Варшавского университета, которая находится по адресу...⁵⁸

Показателен следующий отрывок, в котором можно наблюдать чередование основных содержательных компонентов туристического дискурса — нейтральной дескрипции (в настоящем времени),

⁵⁵ Греция: путеводитель (серия «Вокруг света»). М., 2011. С. 56–57.

⁵⁶ Польша: путеводитель (серия ORANGEвый гид). М., 2013. С. 76.

⁵⁷ Там же. С. 78.

⁵⁸ Там же. С. 85.

нарратива об истории создания памятника (в прошедшем времени) и актуального нарратива-маршрута (снова в настоящем):

*В юго-восточной части Рыночной площади (Rynek Główny) **находится** небольшой храм. Это костел Святого Войцеха — первого святого Польши. Он **был построен** на месте деревянного храма в XII в. Интересно, что костел **находится** на 2,5 м ниже уровня земли.*

*Далее **мы идем** к одиноко возвышающейся Ратушной башне. Башня — это то, что осталось от готической ратуши XIV в.⁵⁹*

РАЗГРАНИЧЕНИЕ ДИКТУМА И МОДУСА

Для дальнейшего изложения методики коммуникативного анализа текста следует обсудить комплекс вопросов, связанных с оппозицией «диктум — модус», а именно: разграничение пропозиций факта и мнения, определение субъекта мнения, анализ оценочного компонента суждений и его отражение в информационной структуре текста, учет коннотаций лексических единиц и др.

Содержание языковых средств, используемых автором для воплощения сведений об известных ему фактах, своей трактовке событий, отношении к ситуации, о качестве взаимоотношений между людьми и т. д., состоит из двух различных компонентов: диктумной и модусной части⁶⁰. Каждый из этих понятийно-семантических компонентов содержания языковых средств характеризует принципиально разные стороны отраженной действительности.

Диктумная часть (или диктум) — это верифицируемый (рационально-когнитивный) компонент языковых средств, который составляют семантические признаки, характеризующие объективное, конкретное, предметное знание, основанное на эмпирическом, ра-

⁵⁹ Польша: путеводитель. С. 128.

⁶⁰ В отличие от теории высказывания Ш. Балли, в которой термины средневековой схоластики «диктум» (лат. *dictum* — слово, выражение) и «модус» (лат. *modus* — способ) используются для обозначения объективной и субъективной частей суждения (предложения), в данной методике этими терминами обозначаются частные компоненты семантической функции различных видов языковых средств (морфологических, синтаксических, лексико-семантических, функционально-прагматических). Информация о рациональном содержании языкового средства именуется диктумом, а свойственное определенной части языковых средств выражение субъективно-личного отношения к рациональному содержанию — модусом.

циональном освоении результатов непосредственного воздействия на органы чувств человека вещных, материальных свойств предмета. Диктумом охватываются материально-эмпирические, условно-фактические признаки понятий и утверждений, отражающие в сознании объективные свойства и качества денотатов (т.е. внеязыковых референтов, понятия о которых обладают категорией вещиности). Сюда входят понятия, обобщающие чувственные образы воздействия внешней силы на материально-вещественные объекты, включая человека, понятия, характеризующие природные процессы, конкретные физические действия человека. Диктумные признаки понятий связываются в суждениях на основании природной смежности, совместимости, равносущности их реальных денотатов на оси «целое — часть» фрагмента действительности, конкретизируя «часть» реальности. Диктумная часть содержания языковых средств осознается как фактическое (констатирующее, объективное) знание о материально-вещественных объектах, об их количественных характеристиках (по параметру «мало — много»), внешних объективно-качественных свойствах и предикатных способах их пространственной и временной реализации (по параметрам «далеко — близко», «быстро — медленно», «долго — кратко», «часто — редко», «начало — конец»).

Модусная часть (или модус) — это интерпретирующий (эмоционально-прагматический) компонент языковых средств, который включает семантические и стилистические признаки, характеризующие субъективно-качественные и экспрессивно-оценочные аспекты значения высказывания. Модусную часть содержания составляют значения понятийных признаков, отражающих субъективно-ментальную, оценочную характеристику объектов высказывания. С помощью модуса характеризуется степень значимости (оценка важности, ценности, полезности), целесообразности или необходимости предмета речи для говорящего. Модус объединяет эмоционально-экспрессивную, регулятивную часть признаков понятий и суждений, которые связаны субъективными конструктивно-логическими обобщениями.

В модусной части фиксируется опыт оценочной интерпретации качественных (волевых, интеллектуальных, моральных, нравственных, эмоциональных, идеологических, этических) сторон деятельности одушевленных существ, стихийных проявлений сил природы и природных процессов. Модусные части суждений с признаками субъективно-качественной характеристики свойств объектов и про-

явлений предикатных признаков образуют мыслительные структуры, отражающие результаты личного восприятия качественных характеристик материально-вещественных объектов.

Модусы выражают оценки лица, его деятельности, а также оценки поступков физического лица с точки зрения здравого смысла, морали или с правовой точки зрения (в той мере, в которой это может понимать любой дееспособный гражданин, не имеющий специальных познаний в области юриспруденции). Такие понятийные признаки связаны имплицативными отношениями, и в суждениях о денотатах осознаются как выводы из фрагмента действительности на оси «общее — частное», наслаиваясь при этом на диктумную часть значения как «частное».

Модусная часть содержания языковых средств оказывает регулятивное речевое воздействие на мнение и сознание слушателей, поскольку обладает свойствами субъективной информации (с параметрами «плохой — хороший», «справедливый — несправедливый», «честный — нечестный», «объективный — субъективный» и т.п.). В частности, укажем на значения с негативным модусом, интерпретирующим диктумную часть лексического содержания языкового знака: *сговор, шельмовать, клеветать, всучить, навязать, самоуправство, пресловутый, вопиющий, зачинщик, пособник, подлец, фанатик, экстремист, военищина, бодрячество, потуги, интервенция, сборище, произвол, водяра, винище, морда* — или на модусную часть значения словосочетаний: *нелепые бредни, подлая демагогия, пустые фантазии, тоталитарная секта, беспробудный пьяница, усатая гадина* и т.п.

Продолжая иллюстрации диктумных и модусных языковых средств, процитируем слова П. А. Столыпина: *«Народ, не имеющий национального самосознания, — есть навоз, на котором произрастают другие народы»*. Здесь используются значения слов с диктумным компонентом, которые участвуют в конструкции образного сравнения (*народ есть навоз*). Само сравнение, правда, содержит негативную характеристику субъекта сравнения. Приведем также высказывание А. В. Суворова: *«Я русский, — какой восторг!»,* состоящее из диктумной (*я русский*) и модусной (*какой восторг!*) части⁶¹.

⁶¹ В последнее время обе цитаты используются в национально ориентированном дискурсе. См., например: URL: http://www.soprotivlenie.su/slovo/ya_russkiy_kakoy_vostorg.html (дата обращения: 15.03.2018); URL: <http://slavyanskiysoyuz.ru/26-my-russkie-kakoy-vostorg.html> (дата обращения: 15.03.2018); URL: [http://gidepark.ru/user/875461648/content/653617\\$](http://gidepark.ru/user/875461648/content/653617$) (дата обращения: 15.03.2018); URL: <http://rrusia.livejournal.com/5980.html> (дата обращения: 15.03.2018) и мн. др.

Выбор языковых средств с определенным диктумом и модусом осуществляется автором и зависит от качеств темы сообщения и отношения автора к предмету речевой характеристики. Диктумный аспект языковых средств обеспечивает указание на объективные признаки предмета речи, характеризующие внеязыковой референт, поэтому общность диктумных компонентов определяется главным объектом речи и темой сообщения. Языковые средства с нейтральным модусом широко используются в сообщениях о ситуациях, событиях, явлениях, о лице, поступках людей, которые обычно не описываются в ценностных категориях (к такой информации неприменимы оценки «хорошо» и «плохо»).

Если диктумом обладают все инструментальные и коммуникативные языковые средства, то модусом (положительным или отрицательным) обладают далеко не все из них. Именно значения семантических признаков диктума создают когеренцию языковых средств, а вот использование в речи средств с модусным компонентом целиком зависит от воли автора и не влияет на порядок и правила понятийно-семантического связывания частей высказывания. Увеличивая количество языковых средств, обладающих модусом, автор усиливает речевое воздействие на эмоциональную сторону сознания слушателей, поскольку модусный аспект языковых инструментов обеспечивает выражение авторского отношения к предмету мысли и его признакам. Все модусные коннотации должны обязательно указываться в словарном описании значений слов, которое является ядром используемых нами процедур коммуникативно-понятийного анализа текста.

Глава 5

ФУНКЦИОНАЛЬНО-ПРАГМАТИЧЕСКИЙ АНАЛИЗ ТЕКСТА

ВЫЯВЛЕНИЕ ПРОПОЗИЦИЙ И ИХ АНАЛИЗ

Цель аналитических процедур, которые выполняются в автоматическом режиме, состоит в обнаружении и обосновании связи между средствами плана выражения с элементами плана содержания текста. При переходе от анализа грамматических значений, выражаемых формой слова к значениям, вычисляемым путем анализа морфосинтаксических и лексико-семантических связей компонентов отдельного предложения и далее с помощью функционально-прагматического анализа понятийных связей фрагмента текста, важно соблюдать непрерывность лингвистического описания, вскрывать и контролировать содержание понятийных отношений, возникающих между пропозициями (минимальными предикативными единицами — наследниками элементов синтаксических конструкций). В этой связи морфологическое исследование текста является необходимой, но частной задачей, подсобным, вспомогательным видом изучения информационных материалов. Результаты этого традиционного — школьного — вида анализа текста позволяют установить компоненты синтаксической структуры предложения. А те, в свою очередь, дают возможность определить значение конкретных лексико-грамматических отношений как внутри словосочетаний (сегментов предложений), так и между ними.

Если подлежащий автоматическому анализу текст содержит всякого рода неправильности, то его необходимо нормализовать, т.е. привести к такому виду, чтобы сам текст и результаты его анализа были понятны (хотя бы человеку). Для этого написание слов, порядок следования синтаксем и синтаксических конструкций следует привести в соответствие с действующими правилами орфографии, пунктуации и словоупотребления. Поскольку нормализация текста является одним из элементов его интерпретации, следует внимательно следить за порядком и последовательностью замены неверного употребленного или непонятого языкового средства на

верное, понятное, соответствующее языковым правилам и нормативному употреблению.

Чтобы объяснить функционально-прагматические свойства данного словесного знака, нужно его интерпретировать. Интерпретация знака — это операция установления понятийного равенства, выполняемая путем замены исходного знака другим знаком или набором знаков на основе общности значений признаков мыслительного содержания знаков, смыслового соответствия одного знака другому. Например, в справочниках по языку (словарях, грамматиках) понятийное соответствие указывается с помощью стилистических помет, толкования слова, указания синонимических соответствий слов, участия в синтаксических конструкциях и т. п. Однако в случаях, содержащих явления грамматической омонимии, такая характеристика становится недостаточной. Для определения значения слова или синтаксической конструкции необходимо использовать фрагмент текста, представляющий собой полное высказывание о предмете речи, в котором содержатся понятийные признаки, необходимые для восстановления свернутых риторических пропозиций и установления значений межпропозиционных связей. Обратимся снова к предложению *У пострадавшего от нападения балетмейстера Филина улучшается зрение*¹. Из анализа предложения невозможно однозначно понять синтаксическую роль фрагмента *от нападения балетмейстера Филина*. С равными формальными основаниями его можно рассматривать как предложно-падежное словосочетание, имеющее либо обстоятельственно-атрибутивную связь с причастием *пострадавший* (пострадавшего от нападения балетмейстера Филина), либо номинативно-атрибутивную связь адъективированного прилагательного *пострадавший* (осложненной предложной формой субстантива) с именем собственным *Филин* (пострадавший от нападения балетмейстер Филин). В результате мы получим как минимум два варианта синтаксического анализа предложения, только один из которых может быть подтвержден данными из более широкого контекста.

При анализе отдельных предложений следует учитывать значения связей всех компонентов, составляющих синтаксическую структуру предложения. Например, в предложении: *Одно и то же лицо не может занимать должность Президента Российской Федерации*

¹ Заголовок публикации см. в Интернете (URL: <http://www.newsru.com/cinema/29jan2013/filin.html> (дата обращения: 11.04.2018)).

более двух сроков подряд² словосочетание *более двух сроков подряд* использовано в функции обстоятельства временной длительности к сказуемому *занимать*. Наречие *подряд*³, являющееся зависимым словом при сравнительном именном словосочетании *более двух сроков*, характеризует качественный признак временной длительности: 'без перерыва, в продолжение какого-л. отрезка времени'. Обстоятельство *более двух сроков подряд* ограничивает временную сферу действия составного глагольного сказуемого *не может занимать* следующим образом: 'занимать → *как долго?* = более двух сроков, → *идущих каким образом?* = без перерыва'.

Интерпретация ограничения временной сферы в смысле 'занимать более двух сроков, в том числе и подряд' противоречит реальному содержанию предложения, поскольку при такой экспликации содержания будут неверно истолкованы коммуникативные связи наречия, а интерпретация этого ограничения в смысле 'занимать более двух сроков' игнорирует качественные характеристики, вносимые наречием *подряд*. Обе эти интерпретации являются ложными.

Вскрытые конструктивные связи плана языкового выражения и функциональные отношения плана содержания позволяют обосновать лексико-семантические значения главных и второстепенных членов предложения, определить большинство лексических значений словоформ, служебных слов, установить понятийную структуру предложения, выделить риторическое ядро данного предложения и определить связи основной и вспомогательных пропозиций. Для этого необходимо по связанным структурно-грамматическим формам языкового выражения установить лексико-семантическое содержание языковых знаков, выделить диктумную и модусную части лексических значений компонентов расширенной схемы пропозиции, а затем определять ее функционально-понятийный тип.

Речевая структура высказывания, позволяющая различать значения понятийно-риторических единиц (сообщения о фактах, оценочные высказывания, мнения о качествах, отношениях, событиях и т. п.) и связывающая отдельные фразы в смысловое целое, является полипредикативной, так как содержит некоторое количество про-

² См.: Конституция РФ, ст. 81, п. 3.

³ Омонимия наречия и существительного в слове *подряд* здесь не рассматривается, поскольку она снимается на уровне формального синтаксического анализа.

позиций⁴. Она состоит из единиц, минимальный план выражения которых реализуется в простом предложении, а план содержания — в минимальной предикативной единице⁵, равной одной пропозиции. Установление качественных характеристик логико-смысловых связей, существующих у грамматически сопряженных монопропозитивных единиц, является наиболее ответственным этапом функционально-прагматического исследования текста. Только в результате этой процедуры лингвист сможет определить и доказать наличие определенных коммуникативных задач и целевую установку связного текста.

Пропозиция — это минимальная информационная единица содержания текста, которая объединяет в единую структуру когерентные понятийно-семантические признаки предиката и актантов. Количество пропозиций в одном предложении (простом или сложном) может меняться в зависимости от количества предикатов. Количество актантов определяется семантическими валентностями конкретного предиката. Структурным ядром пропозиции является семантический предикат, т.е. понятие, обладающее семантической валентностью на актантные аргументы, а ее понятийным ядром — предметные актанты, обладающие свойством внеязыковой соотнесенности (референтности).

Структура пропозиции представляет собой предикатно-актантный комплекс значений понятийных признаков. Его функционально-семантические характеристики определяются не только значениями категориально-грамматической группы признаков, но и значе-

⁴ Обычно термин *пропозиция* используется для обозначения некоей семантической константы (диктума, интенционала, когнитивной части предложения), которая противопоставлена субъективной переменной, характеризующей модальные и прагматические компоненты предложения. При таком понимании пропозиция соответствует логическому внеязыковому факту, содержащемуся в предложении. В настоящей книге термин «пропозиция» понимается шире — это являющаяся частью высказывания целостная предикатно-актантная единица, в которой сохраняются диктумные и модусные значения, характеризующие данную языковую форму, причем диктумное значение формирует основную часть понятийно-риторической пропозиции, а модусное значение — ее дополнительную часть.

⁵ Минимальная предикативная единица характеризуется когерентными связями (предикатной значимостью) и обладает несколькими свойствами: она способна к распространению и осложнению, она обладает субъектным и предикатным семантическими компонентами, грамматически отнесенными к определенному общему пространственно-временному плану.

ниями группы категориально-понятийных признаков. Выделение первой группы признаков значений позволяет по форме синтаксического предиката (сказуемого) и его актантов (подлежащего, дополнения, обстоятельства) установить количественный состав пропозиции. Использование второй группы признаков позволяет по когерентным значениям понятийных признаков⁶ подтвердить лексико-семантическое значение предиката (предикативного понятия) и его актантов.

Важной частью установления функционально-понятийного типа пропозиции является семасиологический анализ и характеристика языковых средств, использованных в предложении. С их помощью выделяются категориально-понятийные признаки значений лексических единиц, составляющих формально-грамматическую структуру предложения, затем определяется структура пропозиции и свернутые пропозиции воссоздаются до полного простого предложения. Данные о понятийных признаках, реализованных в значении слова, должны содержаться в понятийно-семантическом словаре, который является неотъемлемой частью разрабатываемого нами лингвистического процессора. Если требуемое значение лексической единицы в словарях отсутствует или ее толкование не совпадает с употреблением в конкретном контексте, необходимо давать самостоятельную интерпретацию значения данного речевого употребления слова в соответствии с разработанными для нашего словаря способами.

Таким образом, функционально-прагматический анализ предложения состоит из установления синтаксической когезии связанных элементов предложения, определения понятийной когерентности языковых средств, выделенных по значениям их семантических признаков, и выделения пропозиций (основных, дополнительных, свернутых). В зависимости от грамматической формы предиката, его функциональных и понятийных отношений с актантами выделяется четыре основных типа риторических пропозиций: факты, мнения, оценки и волеизъявления. В свою очередь? эти функционально-понятийные типы различаются по значениям в зависимости от семантических характеристик понятийного субъекта, его предикативного признака и объекта.

⁶ Например, в заголовке статьи *Польский депутат пообещал «раскурить косяк» в здании парламента* значение омонимичной лексики *косяк* можно определить и обосновать когерентностью значений признаков у слов *раскурить* и *косяк*.

Задачей функционально-прагматического анализа является последовательное преобразование формально-грамматической структуры предложений в пропозиции с такими понятийными признаками, которые позволят при синтезе содержания композитного высказывания сопрягать его риторические части в смысловое целое.

Цель функционально-прагматического исследования состоит в определении понятийных функций пропозиций по когерентным признакам их компонентов. Этот вид анализа текста является важным этапом работы, поскольку от его результатов зависит обоснованность выводов о понятийно-риторической структуре композитного высказывания.

Рассмотрим последовательность выделения свернутых пропозиций, входящих в состав простого предложения⁷. Например, понятийно-семантический анализ предложения *На Сахалине возмущены заявлением главы МИД РФ о мирном договоре с Японией*⁸ позволяет установить, что его прагматическая структура состоит из трех пропозиций:

- 1) 'люди, живущие на Сахалине, испытывает чувство возмущения' (первый актант — восстанавливается по значению ментального глагола *возмутить* и значению субстантива *заявление* в творительном падеже, мотивированного информационно-речевым глаголом совершенного вида *заявить*);
- 2) 'глава МИД РФ сделал заявление, касающееся мирного договора с Японией' (интерпретация пропозиции, свернутой в субстантиве *заявление*);
- 3) 'заявление главы МИД РФ стало причиной возмущения людей, живущих на Сахалине' (пропозиция, восстановленная из предикатного синтаксического отношения «глагол с причинной валентностью на творительный падеж субстантива»)⁹.

⁷ Функционально-коммуникативный вид анализа предложения, предлагаемый в настоящей работе, опирается на результаты исследований и опыт описаний, изложенный в монографиях [Белошапкина, Брызгунова, Земская и др. 1989; Золотова, Онипенко, Сидорова 2004; Стародумова 2005].

⁸ Заголовок статьи, опубликованной на сайте: URL: <http://bre.ru/news/22763.html> (дата обращения: 14.04.2018).

⁹ Из этого предложения можно также получить импликацию «существуют некие социально-политические проблемы, связанные с мирным договором между РФ и Японией» (бытийная импликация всего предложения).

Из всех выделенных пропозиций, только вторая пропозиция имеет лексико-семантические признаки факта. Пропозиции 1 и 3 являются фактоидами¹⁰, определяемыми по ментальному значению признака 'сфера действия' их предикатов, а пропозиция 1 является фактоидом также и по значению неопределенного обобщения семантического признака 'человек' первого актанта (пациенса).

Прагматическая структура этого предложения включает все выявленные пропозиции: 'люди, живущие на Сахалине, испытывают чувство возмущения, причиной которого является заявление, сделанное главой МИД РФ, касающееся мирного договора РФ с Японией'.

При функционально-прагматическом анализе важно различать сообщение, имеющее языковую форму факта, и действительное событие, внеязыковую ситуацию. Все то, что реально происходит или произошло в мире, истинно¹¹. Объективный мир, существующий вне нашего сознания, будучи освоенным и познанным, приобретает свойство взаимосвязанных фактов. Однако понятийную форму факта реальность обретает только в сознании. Факты производятся нашим сознанием, вне сознания они не существуют, так же как в нашем сознании не существует никакой иной реальности, кроме той, что была воспринята и освоена нашими органами чувств. Любой эмпирический опыт, являясь отражением и характеристикой фрагмента реального мира, является при этом элементом сознания. Любые новые факты осмысливаются и осваиваются нашим сознанием, объединяясь, встраиваясь и наслаиваясь на схожие факты прошлого опыта. Факт есть элемент мыслимой реальности, которая отличается от мира, существующего вне сознания. Ведь мир, отраженный и хранящийся в сознании, и мир, пребывающий вне сознания, представляют собой принципиально разные сущности, отличные по своей природе.

Наличный текст есть объективный факт — часть реального мира. Такие информационные факты, являясь объективными по форме, субъективны по содержанию. Они имеют речевую природу и отражают мыслительное содержание, характеризующее как объ-

¹⁰ Обобщенные суждения (с квантором всеобщности), в которых характеристики объектов речи могут быть верифицированы лишь частично, поскольку являются одной из разновидностей мнения.

¹¹ Акты коммуникации также относятся к истинным событиям и являются информационными фактами. Именно информационные факты подвергаются анализу слушателей и исследователей с целью извлечения смысла.

ект мысли, так и сознание говорящего. Если автор анализируемого материала воспроизводит чужой текст, то он пользуется цитатой как объективным фактом, не зависящим от воли и сознания автора.

Лингвист должен уметь выделять и различать объективные и субъективные аспекты знания в содержании информационного факта. Объективное знание — это стабильное консервативное знание о физических свойствах и количественной стороне материальных предметов. Оно не зависит от закономерных или случайных изменений внешнего мира, поскольку, отражая фундаментальные законы природы и воплощая опыт многих предшественников, подтверждается частным опытом, осознаваемым как факт. Объективное знание позволяет сделать закономерные прикладные выводы дедуктивным путем. В частности, объективация информационного факта осуществляется указанием на первоисточник, содержащий этот факт.

Субъективное знание — это личное знание об идеальных свойствах качественных сторон материальных предметов, получаемое индуктивным путем в результате обнаружения сходства объектов. Оно абстрагировано от значительной части конкретных условий реального мира, поэтому носит вероятностный характер и способно изменяться вместе с накоплением опыта наблюдения, увеличением объектов рассмотрения и их признаков. Субъективное знание позволяет сделать общие (теоретические) выводы, поскольку оно образуется в результате использования обобщающих качественных характеристик схожих объектов и установления зависимостей между ними на основе интуиции. Субъективное знание является таким же концептуальным, как и объективное знание, но оно имеет статус возможно истинного или гипотетического знания, осознаваемого как мнение (частное или общее).

В практике речевой деятельности субъективное и объективное знание легко сочетаются — любой автор, описывающий действительные события, интерпретирует свои впечатления, согласовывая их со своим знанием мира. Выделяя интегральные и дифференциальные понятийные признаки явления или события, говорящий обобщает свои впечатления не только на основании своего жизненного и языкового опыта, но сообразуясь также с целями и задачами сообщения. Суждения о фактах выражаются и реализуются в речи при помощи лексических средств с предметно-вещественными, конкретно-материальными значениями их интегральных семантических признаков. Поскольку пропозиция факта должна обладать

всеми атрибутами объективного мира, то в предложении, обозначающем факт, семантические признаки главных и второстепенных членов предложения должны быть когерентны по значениям предметно-вещественных, конкретно-материальных признаков.

С точки зрения отражения в сознании нашего знания и понимания действительности понятие факта формирует одну из опорных понятийных категорий высказывания. Из фактов составляется мыслимая картина мира, факты позволяют прогнозировать будущее и объяснять произошедшее. Любой текст состоит из фактов и различных комментариев к ним. При анализе содержания текста следует различать, во-первых, сообщение о факте и комментарии по поводу факта, во-вторых, способы риторического использования фактов в описании, повествовании, рассуждении.

Сведения, излагаемые в форме факта, слушатель может верифицировать, т. е. мысленно соотнести понятийные признаки сказанного с теми признаками, которые он сам может выделить в действительных событиях объективными способами. В результате верификации речевые факты могут оказаться истинными (соответствовать действительности), ложными (не соответствовать действительности) или правдоподобными (частично истинными, частично ложными).

Частично истинные сведения будем называть фактоидами (они рассматриваются в п. «Мнения»). Фактоиды содержат как объективные, так и субъективные элементы знания. Они верифицируются косвенно путем выделения той части информации, которая может быть проверена методами наблюдения, измерения или подтверждения экспериментально, и той ее части, которая отражает результаты интерпретирующего обобщения действительности.

Сведения, излагаемые в форме оценки, слушатель может только интерпретировать, т. е. связать понятийные признаки, характеризующие содержание оценки, с качественными признаками объекта оценки. Пропозиция оценки соотносится с действительностью только в части наличия самого объекта оценки, а в части качественных признаков, придаваемых объекту оценки, слушающий доверяет (или не доверяет) говорящему, поскольку сама оценка зависит от точки зрения говорящего, от его отношения к фактам, мотивировавшим данную оценку.

ФУНКЦИОНАЛЬНО-ПОНЯТИЙНЫЕ ТИПЫ ПРОПОЗИЦИЙ

В аспекте прагматических функций высказывания содержащаяся в нем информация может быть сведена к четырем типам речевых структур, отражающих особенности восприятия действительности говорящим. Это факт, мнение, оценка и волеизъявление. Эти функционально связанные комплексы, воплощаемые в речевых действиях, мотивируют подбор структурно-информационных единиц, формирующих композитное высказывание. Они существенно различаются с точки зрения способов отражения и преломления сведений об объектах реального мира, хотя основаны всего на двух характерных способностях человеческого сознания: способности сознания быть направленным и постигать какой-либо реальный объект вне себя (интенциональность мышления) и способности постигать самое себя и свое внутреннее состояние, чувства, эмоции (рефлексивность мышления). В дальнейшем рассмотрении мы выделим познавательные критерии определения типа речевой структуры по особенностям содержащейся в нем информации, отражающей и преломляющей объекты внеязыковой действительности.

Факт

Факт есть результат отражения и переработки данных чувственного опыта. Он соответствует рационально-когнитивному, эпистемическому типу информации. Факты — это констатирующие утверждения о связи объектов и явлений действительности, сообщения о данных опыта, практики, участниками которых являются одушевленные и неодушевленные объекты, сведения о количественных и качественных характеристиках объектов, воспринимаемых органами чувств, или их пространственно-временных свойствах. В этом рационально-эмпирическом типе информации закономерные характеристики предметов внешнего мира согласованы со знанием законов природы. И хотя факты, наряду с любыми другими сведениями, существуют только в сознании, их можно подвергнуть процедуре верификации — проверить соответствие сведений, изложенных в форме факта, действительному положению дел. При этом ответственность за истинность или ложность факта лежит на авторе сообщения. По сути дела, любой факт есть суждение об истинности или ложности некоего результата, поэтому факт должен подтверждаться (верифицироваться) чувственным, эмпирическим опытом

восприятия и воздействия на сознание человека проявляемых признаков предметно-вещественных объектов. Факты формируют ту часть содержания высказывания, которая представляет в сознании (репрезентирует) знание о конкретных явлениях действительности. В отличие от других сведений, факты — это реально свершившиеся или закономерно повторяющиеся события, познанные характеристики объектов, обладающих свойствами материальности, их отношения и зависимости, существующие в природе.

При анализе речевой практики сообщение о факте можно рассматривать в онтологическом (не зависящем от сознания) и логико-гносеологическом планах. В плане онтологии фактами являются любые не зависящие от наблюдателя состояния действительности или свершившиеся события. Такие факты проверяются инструментальными средствами технических экспертиз.

В логико-познавательном плане фактами называют обоснованное знание, полученное путем описания отдельных фрагментов реальной действительности, существующих в известном пространственном и временном континууме. При таком подходе фактами признаются элементы системы знаний, которые, будучи выражены средствами естественного языка, воплощают онтологические факты. Они формируют ту часть общественного знания, которое представляет (репрезентирует) в сознании опыт познания конкретных явлений действительности.

Следует помнить, что верификация факта является специальной процедурой, выходящей за пределы лингвистического исследования. Лингвист обязан установить наличие в высказывании констатирующих утверждений, которые имеют языковые свойства факта, а подтвердить их соответствие действительности должен специалист в той области, к ведению которой относятся эти факты (исторические, социальные, религиозоведческие, психологические, политические и т. п.).

Способность факта к верификации не отменяет возможность использования в ней слов с оценочным модусом. В этом случае констатация фактического события дополняется сведениями об отношении говорящего к событию. Оно может быть негативными, нейтральными, позитивными.

План выражения глагольных форм, способных манифестировать пропозицию факта, ограничен личными формами глаголов несовершенного вида настоящего и прошедшего времени в изъявительном наклонении, а также личными формами глаголов совершенного вида прошедшего времени в изъявительном наклонении.

Любому факту свойственна эмпирическая и дейктическая определенность. Факт обозначается, поскольку он имеет денотативные и референтные свойства, а прагматическая пропозиция факта выражается в форме утверждения — высказывании, в котором отображается понятийная связь конкретного предмета и его апостериорных признаков. Утверждение грамматически выражается формой повествовательного предложения — как невосклицательного, так и восклицательного. Утверждение может дополняться словами и словосочетаниями, подчеркивающими достоверность сообщаемого (например: *известно, точно, доподлинно, без сомнения*).

Факт — характеризующее объект мысли утверждение о событии, происшествии или явлении, которое с точки зрения языкового выражения воспринимается как существовавшее или существующее на самом деле, в действительности. Утверждение о событии как о факте возможно лишь в том случае, если утверждение касается конкретных объектов или количественно ограниченных их совокупностей.

Утверждение в пропозиции факта всегда конкретно. При этом подчеркнем, что характеристика высказывания «быть утверждением о факте» — это свойство самого высказывания, форма выражения информации, объективно ему присущая. Утверждение не перестает быть утверждением, даже если ему предшествует информация о том, что текст отражает только личное мнение автора. Любое высказывание отражает результаты личного восприятия автора. Однако они могут быть выражены как в форме предположения, так и в форме утверждения.

С помощью фактов сообщается об изменениях, которые происходят или произошли в действительности, о наличии предметно-вещественных объектов и их физических свойствах, о состоявшихся событиях в обществе. Причиняющей силой таких изменений обладают одушевленные объекты и силы природы. Значения глаголов с интегральным признаком физического действия, воздействия на физические свойства объекта, деятельности, физиологического состояния, восприятия физических свойств предмета, его нахождения в пространстве, существования, именованя, перформативности участвуют в качестве предикатного ядра пропозиции факта. В пропозициях с неодушевленными синтаксическими субъектами заняты значения глаголов с интегральным признаком процесса, функции, местоположения, наличного бытия. Связь между условиями и событиями важна при анализе причинно-следственных результатов, в которых предикатный признак всегда имеет соответствующего

качества причину, способную его произвести. Для действий — это живое существо, для процессов — силы природы.

Материально-вещественные денотаты живой и неживой природы образуют онтологические классы семантических актантов, формирующих материальную основу пропозиции фактов:

- физических фактов: действие, воздействие, процесс, положение, состояние денотатов живой и неживой природы;
- логических фактов: констатация бытийности, верификация чувственно воспринимаемых свойств, родо-видовая идентификация денотатов живой и неживой природы;
- количественных данных: количественные характеристики объектов живой и неживой природы;
- социальных фактов: результаты конкретной чувственно воспринимаемой деятельности людей, которые могут быть обобщены и выражены количественными характеристиками.

Категориально-понятийные признаки факта ограничены семантической функцией предикатной репрезентации значений признаков синтаксических актантов, присущих глагольным значениям. В предикатно-актантном комплексе факта преобладает диктумный компонент понятийного содержания, поскольку факты выражают объективное, конкретное, предметное знание о понятиях, материально-вещественные аналоги которых подчиняются законам бытия и не зависят от пристрастий и интересов человека. В сфере фактов используются материально-эмпирические признаки актантных понятий, которые формируют утверждения об объектах, связанных предметно-логическими отношениями.

Факт — это пропозиция, в которой предикативный признак устанавливает или отрицает наличие эмпирически проверяемой связи между конкретными носителями материально-вещественных признаков понятий, между предметным понятием и его количественным или материально-физическим признаком. Пропозицию факта составляет предикативное понятие, актантные места которого заполнены понятийными единицами с денотатными категориальными свойствами (т.е. такие, которые имеют признаки пространственной отграниченности, локализованности во времени и пространстве, референтности физическим телам). Актанты пропозиции факта заполняются значениями существительных с онтологическими признаками:

- значения существительных с репрезентативной функцией названия денотатов (имена собственные);

- значения существительных с материально-вещественными контенсиональными¹² понятийными признаками денотата (нарицательные существительные, способные совмещать функцию описания предмета с его речевой репрезентацией);
- значения существительных с идентифицируемыми признаками денотатного понятия (временные, пространственные, чувственно, интеллектуально познаваемые признаки материально-вещественных понятий).

Отличительной характеристикой факта является его свойство выражать проверяемое, объективное знание о мире, которое может быть либо истинным, либо ложным. Поскольку знания о мире согласуются с общей парадигмой законов бытия, то они подчиняются этим законам и проверяются на соответствие им. В широком смысле пропозиции факта выражают апостериорные знания, полученные людьми при взаимодействии с природой, с материально-вещественными объектами.

Рассмотрим следующий пример: *Собрание длилось два часа*. В этом утверждении субстантив *собрание* имеет значение 'совместное присутствие членов какой-либо организации, коллектива для обсуждения, решения каких-либо вопросов' и представляет собой свернутую пропозицию. В данном случае наличествуют две субъектно-характеризующие пропозиции факта с целевой связью: {люди собрались в одном месте в одно время} 'для того, чтобы' {эти люди обсудили, решили какие-либо определенные вопросы в определенном месте в течение определенного времени}¹³. Поскольку эти свернутые пропозиции имеют одни те же актанты (неопределенный субъект, неопределенный локатив и определенный темпоратив), а их предикатные признаки объединены каузативно-целевой связью, то мы можем преобразовать две пропозиции в одну — субъектно-характеризующую целевую пропозицию: 'неопределенные люди собрались с целью обсуждения, решения некоторых вопросов в течение определенного времени и в неопределенном месте'. Эта пропозиция,

¹² Контенсионал, или содержание понятия о предмете — совокупность, устойчивая структура отраженных в данном понятии (значении имени) материально-вещественных признаков. В этом случае термин «интенционал» понимается как «стабильное ядро, устойчивая часть вариативного контенсионала».

¹³ Целевая связь конкретизирует экстенционал актантов первой пропозиции, поэтому во второй пропозиции актанты становятся определенными (т. е. те самые люди, которые там и тогда собрались, они же и обсуждали что-то там и тогда).

характеризующая субъектное понятие *члены какой-либо организации, коллектива*, осложненное предикатным конкретизатором (люди, обсуждающие проблему), имеет предикативный признак, устанавливающий ее временные границы — *длился два часа*. В результате функционально-прагматического анализа этого предложения должно быть констатировано сообщение о факте: собравшиеся люди обсуждали какие-то вопросы в течение двух часов.

Пропозиции фактов можно разделить на несколько видов в зависимости от значений признаков предикативных понятий: физические факты (действия, поступки, деятельности конкретных людей, природные процессы с определенным пространственно-временным континуумом), социальные факты (события, ситуации, не зависящие от воли одного человека, случившиеся в определенном пространственно-временном континууме), бытийные факты (утверждения о существовании конкретного объекта с присущими ему физическими качествами), информационные факты (сообщения о сказанном, оглашенном, решенном каким-либо определенным физическим или юридическим лицом) и количественные данные о материально-вещественных объектах. Возможность проверки фактов на истинность или ложность не исключает того, что использованные языковые средства могут содержать модус оценочного отношения говорящего к сообщаемым сведениям.

При анализе контекста, в котором используется тот или иной факт, необходимо исследовать содержание авторских комментариев, сопровождающих констатацию факта, и характер связи исходного суждения с аргументирующим фактом. В противном случае фрагмент текста, в котором обнародованы статистические сведения¹⁴, без связи с авторскими суждениями не позволит обосновать целевую установку фрагмента. Ср.: *Чаще всего негативные эмоции у сограждан по-прежнему вызывают представители кавказских народов (29 %). Значительно реже говорят о нелюбви к выходцам из Средней Азии (6 %). По 3 % недолюбливают китайцев и евреев. Впрочем, большинство опрошенных одинаково хорошо относится ко всем народам и национальностям (56 %)*¹⁵.

¹⁴ Числовые показатели количественных сведений формируют пропозицию факта (достоверность которого может быть проверена внеязыковыми способами).

¹⁵ Использован фрагмент текста, размещенного на сайте: URL: http://kprf.ru/rus_soc/79465.html (дата обращения: 14.04.2018).

Указание на факт — это всегда указание на частный случай, в котором субъектный и объектный актаны, как и предикатный признак, конкретны. В высказывании он может выполнять различные функции. Эмпирические данные могут использоваться в качестве примеров, иллюстраций и образцов. Факты обычно приводятся в аргументирующей части авторского высказывания или как информационный повод для авторских комментариев.

Если факт указывается в качестве аргументирующего примера, то он используется для подтверждения обобщающего мнения о качествах всего референтного класса. Если факт приводится как пример, подтверждающий реальность какого-либо типа поведения, то он используется как образец (или антиобразец) для подражания.

Если факт приводится как исходное утверждение, а в аргументации он анализируется, обобщается или оценивается, то факт становится темой высказывания, обнаружение факта является задачей высказывания, а комментарии к нему — целевой установкой высказывания.

Рассмотрим пример: *Судебный процесс по делу лидера националистической нидерландской Партии за свободу Хеерта Вилдерса, обвиняемого в разжигании межнациональной розни и дискриминации мусульман, к удивлению многих наблюдателей в Европе, закончился полным оправданием политика.*

Перечислим утверждения о фактах в этом предложении:

- *Судебный процесс по делу Хеерта Вилдерса закончился оправданием политика* (юридический факт — определяем по значению фазового предиката);
- *Хеерт Вилдерс — лидер нидерландской Партии за свободу* (социально-бытийный характеризующий факт);
- *Хеерт Вилдерс обвиняется в разжигании розни и дискриминации мусульман* (юридический факт).

Утверждения о фактах здесь сопровождаются оценочными атрибутами, которыми автор поясняет свою позицию. Например, нидерландскую Партию за свободу он называет «*националистической*», а «*многих*» наблюдателей удивило «*полное*» оправдание политика.

В этом предложении содержится также авторское мнение, которым характеризуется вся фактическая информация («*многих наблюдателей в Европе удивил результат судебного процесса*» — мнение о ментальной ситуации).

Мнение

Мнения представляют собой качественный, аксиологический тип информации. Это этические суждения о ценностных качествах отношений, событий, действий, характеризующие сведения, выражающие оценку человеком качественных связей объектов (явлений, событий и т.п.), основанную на личном восприятии, ощущении, субъективном представлении о значимости (важности, ценности, полезности, моральности, нормативности, нравственности) и, соответственно, о степени влияния объекта суждения на социально-политические, экономические, культурные, интеллектуально-ментальные процессы.

Мнения интерпретируют действительное положение дел путем замены фактической характеристики объекта, которую можно проверить внешним для языка способом, на производную от нее характеристику, зависящую от личных впечатлений субъекта восприятия, от его отношения к объекту речи. Они составляют ту часть содержания высказывания, которая выражает представления говорящего о выявленных и осознанных им признаках ценностных качеств объекта речи (в противопоставлениях типа «правильно — неправильно», «справедливо — несправедливо», «рационально — нерационально»). Мнения формируют качественно-характеризующую часть высказывания, признаваемую справедливой, если они непротиворечиво сопряжены с фактами или знаниями слушателя.

Мнения — это суждения о качественных сторонах отношений, событий, действий, характеризующие сведения, выражающие личное отношение человека к качественным связям объектов (явлений, событий и т.п.), основанное на интуитивном восприятии, ощущении, субъективном представлении о значимости (важности, целесообразности, эффективности, полезности, моральности, нормативности, нравственности, достоверности) и, соответственно, о степени влияния объекта суждения на социально-политические, экономические, культурные, интеллектуально-ментальные процессы.

Мнения не подлежат проверке на истинность или ложность. Они формируют качественно-характеризующую часть индивидуального знания, признаваемого правильным, справедливым, рациональным, если они обобщают достоверные факты и подтверждаются как интуитивными ощущениями, так и собственными обобщающими интерпретациями слушателей.

План выражения глагольных форм, способных манифестировать пропозицию мнения, не имеет ограничений по категориям вида, времени, наклонения. Формально часть пропозиций мнения может быть установлена по наличию в предложении вводных слов и сочетаний с модальным значением неуверенности, сравнения, возможности, предположения, выражающим оценку говорящим степени достоверности сообщения, а также частиц со значением предположения, допущения, возможности, условности, сомнения.

Носителем и выразителем мнения может быть только человек. Выражение мнения предполагает указание на его носителя, поскольку мнение содержит результат интерпретации фактов действительности, результат обобщения личных впечатлений от фактов, а не обобщение самих фактов. Мнение имеет сигнификативные свойства. С помощью мнений сообщается об интеллектуальном, ментальном, этическом, эстетическом отношении человека к фактам, к тем изменениям, которые произошли или происходят в сознании человека под влиянием действий, процессов, событий, произошедших в мире. В отличие от факта, на который достаточно просто указать, авторское мнение требует последующего обоснования, интеллектуальной поддержки. Поддерживающая аргументация предназначена для слушателя, который с помощью мысленного соотношения услышанного с известными ему правилами языка, законами природы и социальными нормами подвергает рациональному анализу отношения между содержанием мнения и аргументацией.

Разница между констатацией факта и мнением о факте состоит в том, что высказанное мнение о факте (событии, лице), в отличие от утверждения, сообщения о факте, предполагает явное указание на носителя мнения. Выражение мнения распознается в тексте по наличию определенных слов и конструкций, указывающих на авторскую интерпретацию действительности с точки зрения важности, целесообразности, эффективности, полезности, моральности, нормативности, нравственности, достоверности. Если такие показатели отсутствуют, то адресат текста (получатель информации) вправе полагать, что имеет дело с утверждением, а не с мнением.

Мнение, в отличие от утверждения о факте, нельзя проверить на истинность или ложность, однако оно может подтверждаться или не подтверждаться фактами, действительными событиями. Мнение содержит информацию, но не о самой действительности (фактах), а о том, какой образ действительности существует в сознании говорящего. Пропозиция мнения отражает личный опыт восприятия

результатов взаимодействия с другими людьми, объясняет причины их чувств, мыслей и поступков на основании навыков саморефлексии. При выражении мнения используются значения предикатных признаков, формирующих модально-ментальные области субъективного знания. Мнение — это всегда оценочное умозаключение, обобщающее ментальные, эмоциональные, этические результаты личных впечатлений.

Категориально-понятийные признаки мнения относятся к семантической области предикатных признаков, которыми обозначаются результаты интерпретации опыта интеллектуальной рефлексии над внешними проявлениями деятельности одушевленных существ (волевой, интеллектуальной, эмоциональной, социальной и т. п.).

В формировании предикативного ядра пропозиции мнения участвуют значения глаголов с интегральным признаком ментально-интеллектуального действия, восприятия эмоционального состояния, сопоставления, сравнения, осознания, понимания, предположения. Понятиям, образующим актантную структуру пропозиции мнения, говорящий присваивает такие свойства и качества, которые связаны между собой отношениями причинности, детерминированности, обусловленности. Мнение встраивается в понятийно-риторическую конструкцию полного высказывания путем создания имплицативных связей с другими субъективными понятиями, с фактами и элементами объективного знания. В широком смысле пропозиции мнения отражают априорные знания человека, получаемые им при освоении социально-культурных традиций здравого смысла, моральных, нравственных норм и эстетических правил, поддерживаемых обществом.

Мнение — это пропозиция, в которой предикатный признак устанавливает или отрицает модально-ментальную связь между носителями материально-вещественных признаков понятий, между качественными характеристиками понятий, между ментальными или интеллектуальными понятиями. Наличие такой связи подтверждается или опровергается путем формально-логического сличения и согласования сказанного с субъективным опытом восприятия действительности. Пропозицию мнения составляют предикативные понятия, актантные места которых могут заполняться не только понятийными единицами с денотатными категориальными свойствами, но и с сигнификатными свойствами (т. е. такими, которые обозначают понятия, не связанные признаками пространственной отграниченности, пространственной локализованности). К актантам с сигнификатными свойствами относятся:

- существительные с характеризующей функцией значений понятийных признаков;
- существительные с экстенционально-предикатными¹⁶ значениями понятийных признаков;
- существительные, обозначающие отвлеченные (нереферентные) понятийные признаки.

Отличительной характеристикой мнения является его свойство выражать субъективное знание о мире, с которым слушатель может согласиться или не согласиться в той мере, в которой это отвечает его частным интересам. Субъективное знание о мире подчиняется законам бытия только частично, оно опосредовано индивидуальным опытом обобщения внешних признаков причинно-следственных связей событий.

Мнение (о фактах, событиях, лицах) — суждение, выражающее чью-то точку зрения, отношение к кому-либо, чему-либо. Оно может быть выражено как в форме предположения, так и в форме утверждения.

Мнение в форме предположения распознается по наличию определенных слов и конструкций (например: *по моему мнению, мне кажется, я считаю, я полагаю, кто-то думает, что, надо полагать, что и др.*).

Мнение в форме утверждения таких маркеров не содержит. Мнение в форме утверждения о состоявшемся событии может быть проверено на полноту соответствия действительности. При этом мнение, в отличие от утверждения о факте, не может быть истинным или ложным, так как отражает не реальную действительность, а ее восприятие человеком, не мир, но его отражение в личной картине мира. Однако оно может подтверждаться или не подтверждаться фактами, событиями объективной действительности. Мнение может опираться на факты, этим оно отличается от домысла, и при этом содержать оценку фактов и их комментарии, в этом его отличие от объективного знания.

Любое мнение содержит информацию двоякого рода: во-первых, констатирующие сведения о действительности — рациональную информацию и, во-вторых, ценностную, этическую информацию. Отображаемая действительность может оцениваться как хорошая

¹⁶ Экстенционал, или объем предикатного понятия, — это совокупность тех значений признаков обозначаемого понятия, которые когерентны значениям признаков актантов, восполняющих данное понятие (значение знака).

или плохая (красивая или уродливая, вредная или полезная, своя или чужая, модная или устаревшая и проч.) в зависимости от той системы ценностей, которой придерживается автор высказывания.

Поскольку в пропозиции мнения субъективная оценка качества объекта высказывания наслаивается на объективную (констатирующую) информацию об объекте, то на соответствие действительности может быть проверено только содержание диктума. В суждении *Обвиняемый X, отказавшись от сделки со следствием, поступил глупо и необдуманно, так как теперь не сможет рассчитывать на смягчение наказания* можно проверить, действительно ли X является обвиняемым, отказался ли он от сотрудничества со следствием и может ли его наказание быть смягчено. Собственно оценочный компонент мнения (*X поступил глупо и необдуманно*) верификации не подлежит, поскольку для одних людей поступок X — плохой, неправильный выбор, а для других — целесообразный, правильный.

В некоторых случаях разграничение утверждений о фактах и выражения мнения следует проводить в зависимости от всего контекста высказывания. Например, фраза X — *фашист* может расцениваться как утверждение о факте, если под этим понимается, что X состоял или состоит в фашистской партии. Если же речь идет об идеологических (политических) взглядах X, то такое высказывание является выражением оценочного мнения.

Мнение может формироваться на основе фактов или же быть предвзятым, т. е. основанным на обобщении намеренно подобранных отдельных фактов. Человек может формировать свое мнение сознательно, рационально оценивая факты, или интуитивно, не анализируя причин, по которым он делает такое умозаключение. В случае когда в тексте изложен ряд фактов, сопровождающийся каким-то выводом, вытекающим из них, то такой вывод может иметь форму предположения. Вывод может быть неверным по двум причинам:

- одно из логически предшествующих ему утверждений о фактах не соответствует действительности;
- ошибочна сама логика рассуждений (факты изложены достоверно, но такой вывод на их основании сделать нельзя).

Мнение может быть выражено в грубой, неприличной форме. Такое оценочное суждение может порочить честь, достоинство человека, позорить его, негативно влияя на деловую репутацию.

Предположение, в отличие от утверждения, отмечается специальными языковыми средствами — словами, выражающими неуве-

ренность, сомнение, вероятность происхождения того или иного события, одну из ряда возможных версий (типа *может быть, вероятно, по-видимому, как представляется, думается*). Предположение является одной из форм выражения мнения, когда автор текста хочет подчеркнуть предварительный характер высказываемых доводов или собственную неуверенность в их достоверности.

Сомнение — разновидность предположения, когда говорящий предполагает, что факта, о котором он сообщает, в действительности не было.

Риторический вопрос — высказывание, вопросительное по структуре, но передающее, как и повествовательное предложение, утверждение о чем-либо. С помощью риторического вопроса говорящий интерпретирует утверждение о чем-либо как самоочевидное, всем известное.

Мнение может быть выражено в отношении ментальных предикатных признаков субъекта пропозиции. Возьмем в качестве примера следующее предложение: *Россия переживает в первую очередь не политический и не экономический кризис, а кризис духовный*¹⁷. В этом предложении утверждается наличие связи между предметным административно-территориальным номинативом (*Россия*) и атрибутированным ментальным субстантивом (*кризис*). Предикатная связь имеет значение ‘подвергнуться чему-л.; испытать какие-л. процессы, явления’. Когерентным признаком, объединяющим пропозицию в понятийное целое, служит агентивный признак со значением ‘человек, люди’, который обнаруживается в метонимическом переносе в слове *Россия* (административно-территориальный объект — люди, населяющие этот объект), наличествует в субstantиве *кризис* и в актантных семантических признаках значения глагольной пары *пережить-переживать*. Субъектом пропозиции являются ‘граждане России’, объектом пропозиции является субстантив ‘резкое изменение духовного состояния людей’. Основная пропозиция этого предложения может быть выражена словами: {все граждане России испытывают резкое изменение своего духовного состояния}. Поскольку автор приписывает определенное ментальное состояние, которое свойственно всем без исключения гражданам России, то такое утверждение следует безусловно расценивать как выражение мнения.

¹⁷ Использован фрагмент текста, размещенного на сайте: URL: <http://www.1-sovetnik.com/articles/article-886.html> (дата обращения: 13.03.2018).

Если субъект пропозиции характеризуется предикативным признаком со значением эмотивного предпочтения, то такая пропозиция также является мнением — мнением о предпочтении (преференции). Например: *Демократы наши любят народ идеальный, отвлеченный, в отношении к которому тем скорее готовы исполнить свой долг, что он никогда не существовал и существовать не будет* (Ф. М. Достоевский. Дневник писателя).

В пропозиции мнения предикативный признак приписывается всему классу материально-вещественных объектов, способных занять позицию актанта (например: *Демократы наши любят народ идеальный, отвлеченный*, т. е. все люди, входящие в класс «наши демократы»), либо нереферентному актанту приписывается предикативный признак как неизменное атрибутивное качество (например: *идеальный, отвлеченный народ никогда не существовал и существовать не будет*).

Мнение может быть выражено по отношению к содержанию ментального субъекта пропозиции в сочетании с мерой и качеством проявления предикативного признака. Ср: *Понятие «национализм» настолько ложно вдолблено в наши головы, что уже неосознанно воспринимается как вызов другим народам, как покушение на их национальный суверенитет, как объявление войны другим народам.*

Выделим основные структурные части данного предложения:

- *понятие настолько ложно вдолблено в наши головы;*
- *что* — изъяснительный союз;
- *понятие уже неосознанно воспринимается как содержащее вызов другим народам;*
- *понятие уже неосознанно воспринимается как содержащее покушение на их национальный суверенитет;*
- *понятие уже неосознанно воспринимается как содержащее объявление войны другим народам.*

Автор выражает свое мнение по поводу содержания понятия «национализм». Он утверждает, во-первых, что кто-то заставил всех людей усвоить понятие «национализм» с «ложными» (т. е. ошибочными, неправильными) признаками, во-вторых, эти признаки воспринимаются как *вызов другим народам, покушение на их национальный суверенитет, объявление войны другим народам.*

В пропозиции мнения часто используются актанты (и партиципанты) с модусом эмоциональной оценки, а также предикатные признаки, содержащие модусный компонент. Например: *И хотя Чаплин*

*придает холодному имморализму и педантизму бывшего счетовода отраженный блеск безупречности каждого движения и жеста Чарли, всякий раз именно этим сходством он проявляет пошлость и механистичность этих имитаций, многих и разнообразных, доброго мужа и пылкого воздыхателя*¹⁸.

Использование в предложении частиц, употребляющихся для обозначения мнимости, несоответствия действительности, для выражения неуверенности, неточности сообщения, частиц для форм сослагательного наклонения, и т. п. переводит содержание пропозиции факта в разряд фактоидов¹⁹ — одной из разновидностей мнения. Фактоиды лишены конкретности, они всегда содержат в себе эмпирическое обобщение. Вместо конкретности очевидного в фактоиде используется узнаваемая наглядность повторяющегося, поэтому в убеждающих высказываниях они часто используются вместо фактов.

Фактоид — мнение, которое воспринимается как факт, но не может оцениваться в категориях «истинно — ложно», а только как верное, вероятное или ошибочное, невероятное суждение. Слушатель вынужден принимать обобщенные качественные признаки, содержащиеся в фактоиде, доверяя компетенции говорящего. В отличие от пропозиции факта, в основе которой лежат предикативные и актантные признаки, характеристики которых могут быть подтверждены средствами объективного контроля, ложность или истинность фактоидов не может быть проверена по объективным языковым причинам — они характеризуются эмпирической (дейктической) неопределенностью. Это позволяет использовать фактоиды для манипулятивного воздействия на мнение человека, поскольку актанты, участвующие в формировании пропозиции фактоида, либо не обладают денотативным значением, либо нереферентны объектам реальной действительности. Если в пропозиции мнения фактические сведения (описание, повествование) преобразуются таким образом, что информация, подлежащая верификации, понятийно интерпретируется, то такие сведения становятся фактоидом:

- **с обобщающими свойствами**, если признаки, характеризующие участников конкретного события, происшествия, обобщаются до уровня признаков, характеризующих всех представите-

¹⁸ Божественный Чарли/Экран и сцена, 2004.05.06. URL: <https://www.facebook.com/screenstage/> (дата обращения: 14.04.2018).

¹⁹ Фактоид от англ. factoid — принимающий вид факта.

лей какой-либо социальной группы, их постоянное поведение, свойства и черты характера и т. п.;

- **с концептуальными свойствами**, если признаки, характеризующие несколько разрозненных событий, происшествий, обобщаются до уровня явлений, имеющих характер социальных и природных закономерностей, получающих логико-аналитическое объяснение в теоретических предположениях (гипотезах).

Факты с референтными актантами, излагаемые в пересказе сплетен, слухов, также становятся фактоидами, поскольку автор намеренно отказывается взять на себя ответственность за достоверность излагаемых конкретных обстоятельств.

Оценка

Пропозиции этого типа выражают экспрессивно-прагматическую, эмотивную информацию. Оценка включает квалифицирующие мнения, выражающие эмоциональную реакцию, личное оценивающее суждение человека по поводу отличительных признаков какого-либо объекта, факта (в явном или неявном сравнении с другим в чем-то похожим объектом, фактом). Оценка составляет ту часть содержания высказывания, которая выражает отношение говорящего к объекту речи в целом, при этом ничего не сообщая о признаках самих предметов.

Оценка как факт индивидуального сознания составляет ту часть вероятностного знания, которое характеризует ценностное отношение говорящего к объекту речи (в противопоставлениях типа «хорошо-плохо», «лучше — хуже», «полезный — бесполезный») и формирует экспрессивно-оценочную часть высказывания, не существующую отдельно от объективных признаков предмета оценки. С помощью оценки — прагматической, эмоциональной, этической, эстетической; нормативной, абсолютной, относительной; позитивной, негативной, критической (рационально обоснованной), некритической (рационально не обоснованной) — устанавливается ценность, значимость какого-либо объекта, события, явления и т. п.

Оценки составляют квалифицирующие мнения, отражающие реакцию и выражающие личное обобщающее суждение человека об отличительных признаках какого-либо объекта, факта (в явном или неявном сравнении с другим в чем-то похожим объектом, фактом) с точки зрения их эффективности, полезности, вредности и пр.

Оценка — это пропозиция, в которой предикативный признак устанавливает или отрицает наличие определительной связи между актантом и ментально-качественным признаком. Оценочное высказывание содержит информацию двоякого рода: во-первых, констатирующую информацию о действительности (диктумную часть) и, во-вторых, интерпретирующую информацию. Любые фрагменты отображаемой действительности могут оцениваться как хорошие или плохие (красивые или уродливые, вредные или полезные, свои или чужие, модные или устаревшие и проч.) в зависимости от той системы ценностей, которой придерживается автор высказывания.

При оценке важен интерпретирующий компонент содержания приписываемого признака, который сообщает слушателю о негативном или позитивном отношении говорящего к объекту высказывания. В пропозиции оценки объекту (предметно-вещественному или предикатному признаку) приписывается качественная характеристика. Тем самым автор присваивает объекту качество, которое уподобляется имманентному признаку.

Для этого автор использует сопоставительное сравнение понятий, т. е. сопоставляются значения признаков двух или нескольких понятий, которые могут быть схожи между собой или совсем не схожи. В основании сравнения могут лежать как общие, так и случайные признаки объектов. Оценка важности или обязательности выделяемых признаков или какой-либо степени схожести объектов всегда является субъективной. Сравнения не подлежат логической верификации, их нельзя проверить на соответствие действительности, поэтому они часто используются в целях манипулятивной аргументации, исподволь воздействующей на сознание слушателей. В значениях предикативных признаков и актантов, участвующих в формировании пропозиции оценки, преобладают модусные компоненты с общим интегральным признаком, расположенным на шкале «хороший — плохой». При оценке признака объекта пропозитивный констатирующий компонент скрыт в бытийных презумпциях (пресуппозициях). Этим пропозиция оценки отличается от мнения.

Рассмотрим следующее предложение: *К сожалению, реальность в Судане говорит, что правительство является марионеточным*²⁰. В данном суждении объекту, соотносимому с понятием «правительство Судана», приписывается признак 'исполняющий чужую волю,

²⁰ Из журнала «Аль-Ваъй» № 231, с. 16.

обслуживающий чуждые интересы' с модусом отрицательной оценки. Презумпциями данного предложения являются следующие компоненты: 'существует государство Судан' и 'в государстве Судан существует правительство'. Отметим, что источником сведений в этом предложении автор объявляет *реальность в Судане*. Этот риторический прием обычно используется для отчуждения авторской оценки, создания впечатления объективности, беспристрастности оценки.

Оценке обычно подвергается не вся понятийно-семантическая структура значений признаков объекта, а только те признаки, которые оказываются связанными данным словосочетанием. Например, определение в словосочетании *острый нож* характеризует не интегральный признак *ножа* (вещь, инструмент), а дифференциальный признак — режущие или колющие свойства его части. В зависимости от объекта оценки можно выделить и разновидности модусных шкал оценок с интегральным признаком «хороший / плохой»:

- рациональные (интеллектуальная) шкалы: «умно — глупо», «целесообразный — нецелесообразный»;
- утилитарные шкалы: «полезно — вредно», «удобный — неудобный», «выгодно — невыгодно», «вкусный — невкусный», «приятно — неприятный»;
- эстетические шкалы: «красивый — безобразный», «архаичный — современный», «оригинальный — вторичный», «хаотичный — гармоничный»;
- социальные шкалы: «свой — чужой», «бедный — богатый», «низы — верхи»;
- этические шкалы: «можно — нельзя», «благородно — аморально», «приличный — неприличный», «тактичный — вызывающий», «справедливый — несправедливый» и т. п.

Каждая из этих шкал может совмещать несколько признаков или разделяться на подвиды. Например, на этических шкалах можно выделить:

- юридическую шкалу: «легальный — нелегальный»; «разрешенный — неразрешенный»;
- политическую шкалу: «прогрессивный — консервативный», «миролюбивый — агрессивный» и т. п.

Такие шкалы субъективных оценок допускают возможность дополнительной модусной характеристики степени проявления признака значениями, выражающими понятия «очень — не очень», «абсолютно — относительно», «предельно — минимально».

Оценки, не содержащие интегральный признак «хороший/плохой», относятся к разряду скалярных количественных характеристик, которые могут быть проверены и подтверждены числовыми значениями. Такие характеристики объектов могут участвовать в пропозиции факта. Например, шкалы приблизительной количественной оценки относительно средней нормы:

- шкалы чувственного восприятия материально-вещественных объектов: «теплый — холодный», «твердый — мягкий», «пресный — соленый», «яркий — темный»;
- количественные шкалы: «много — мало», «высокий — низкий», «быстрый — медленный», «далекий — близкий», «широкий — узкий» и т. п.²¹

Оценка может выражаться в нескольких формах: 1) непосредственная прямая оценка с негативным или позитивным модусом: *Х — плохой (глупый, неприятный; благородный, цветущий) человек*; 2) вероятностная оценка, опосредованная выражением мнения говорящего: *Мне кажется, Х плохой человек; Я думаю, что на них можно положиться.*

Оценка может быть высказана: 1) без прямого сопоставления с другим объектом, ср.: *Деятельность предприятия У приносит пользу горожанам*; 2) в сопоставлении с другими объектами, например: *В отличие от других предприятий города, фабрика У заботится о своих работниках.*

Волеизъявление

Волеизъявления (регулятивно-качественный, деонтический тип информации) формируют регулятивно-управляющую часть содержания высказывания, целью которого является агитационное управление волевыми, осознанными действиями адресата для побуждения его к активной деятельности и достижению целей, устанавливаемых говорящим. Волеизъявление — это речевое воздействие, которое используется для побуждения адресата к активной деятельности для достижения целей, устанавливаемых говорящим.

²¹ То обстоятельство, что для кого-то данная характеристика объекта будет совпадать с личным восприятием, не означает, что пропозиция факта перестает быть таковой. Для нас важно, что автор сделал некоторое утверждение в форме факта. Если бы автор захотел сообщить свое мнение о факте, он указал бы на это, пояснив словами: *по моему мнению, я считаю, мне кажется* и т. п.

К этому типу информации относятся разнообразные приказы, указания, инструкции, наставления, советы, призывы, лозунги, обещания, угрозы, клятвы.

Специфика информации деонтического типа состоит в том, что предикат обязывающего сообщения обозначает потенциально возможное, еще нереализованное событие, действие или процесс. Словесное выражение содержания высказывания говорящего (с различной мерой модальности: «могу — хочу — надо») является коммуникативной задачей, а управляющее воздействие содержанием этого высказывания является целевой установкой информации данного типа. Волеизъявление — это высказывание, в котором говорящий воздействует на адресата формой и содержанием предиката с подчиненными ему компонентами фразы.

Диапазон глагольных форм, способных манифестировать позицию волеизъявления, довольно широк. Он включает средства выражения модальных значений, значений повелительного наклонения, будущего времени. Модальные слова, частицы, выражающие различные модальные значения, участвуют в формировании пропозиции волеизъявления. Главным отличием этого типа пропозиции является фактическая нереализованность его предикатного признака. Говорящий, сообщая о своем отношении к главному объекту высказывания, выражает мнение о необходимости осуществления или неосуществления предикатного признака. Ср.: *Над Россией должен, наконец, воцариться высший интерес, над всем обществом, над всеми политическими дрязгами, и это интерес сохранения, сбережения, приумножения русской нации. Сегодня интересы русского нам должны быть значимее и дороже интересов целых континентов. Сегодня понятие «нация» должно быть нам так же близко, дорого и свято, как понятие «семья»*²².

В волеизъявлении с использованием форм будущего времени говорящий выражает различные оттенки уверенности в реализации предикатного признака. В качестве примера приведем высказывание: *Уважаемые коллеги, давайте будем делать вместе ту работу, от которой зависит благосостояние наших жителей, работать во имя нашей области, во имя успеха!* — сказал глава региона²³.

²² Цитата из сочинений Бориса Миронова, распространяемая в блогосфере и на сайтах Рунета.

²³ URL: <http://seltso-vestnik.ru/society/public-associations/2017/11/17/aleksandr-bogomaz-davajte-rabotat-vo-imya-nashej-oblasti-vo-imya-uspeha/> (дата обращения: 14.04.2018).

Используя инклюзивные формы повелительного наклонения, говорящий побуждает к совместному с ним действию того, к кому обращена речь, ср. *пойдем, пойдете, сделаем, расскажем* в отличие от *пойдите, сделайте, расскажите*.

В качестве языковых форм волеизъявления рассматриваются также регулятивные суждения: призывы, лозунги, девизы, идеологемы. В широком смысле регулятивные пропозиции отражают апостериорный опыт взаимодействия людей, опыт социального управления и подчинения.

Пропозиция волеизъявления является наиболее ясным и понятным типом целенаправленного воздействия на слушателей, если выражается в форме агитационного призыва. Например: *Будьте теплом и светом для окружающих; старайтесь сперва согреть себя семью, трудитесь над этим, а потом эти труды вас так завлекут, что для вас уже узок будет круг семьи, и эти теплые лучи со временем будут захватывать все новых и новых людей и круг, освещаемый вами, будет постепенно все увеличиваться и увеличиваться; так; старайтесь, чтобы ваш светильник ярко горел*²⁴. Подобное высказывание иллюстрирует один из случаев пропагандистского высказывания с элементами агитации, имеющим понятийно-риторические признаки оправдания необходимости осуществления действий, направленных на группу лиц, выделяемую по признаку существования.

КОМПОЗИТНОЕ ВЫСКАЗЫВАНИЕ КАК ЕДИНИЦА РИТОРИЧЕСКОЙ СТРУКТУРЫ ТЕКСТА

Установив типы пропозиций, составляющих фрагмент текста, и определив их связи, можно приступать к синтезу коммуникативно-риторической структуры высказывания. Функциональное описание результатов системной когезии синтаксических единиц позволяет совершить контролируемый выход за пределы отдельного предложения и по когерентным семантическим значениям установить виды контактных и дистантных связей, существующих между пропозициями. Последовательная характеристика таких связей позволяет использовать процедуры связывания выделенных пропозиций по когерентным значениям признаков в единую риторическую конструкцию.

²⁴ URL: <http://www.pravmir.ru/budte-teplom-i-svetom-dlya-okruzhayushhix/> (дата обращения: 14.04.2018).

С помощью выделения в высказывании связанных друг с другом базовых информационных структур (фактов, мнений, оценок, волеизъявлений) наш подход позволяет синтезировать композитное высказывание, которое является единицей риторической структуры текста. Композитное высказывание содержит несколько грамматически выделяемых предложений, каждое из которых в рамках общей функционально-прагматической структуры характеризует один и тот же объект речи, но с разных сторон. Предназначение композитного высказывания состоит в организации общего понятийного пространства характеристик главного риторического объекта, связанных между собой отношениями обусловленности. Единое понятийное пространство — это структура сведений о внешних и внутренних сторонах главного объекта высказывания, объединенная общими пространственно-временными условиями его существования.

Функционально-прагматическое содержание композитного высказывания является базовым результатом автоматического анализа текста. Это тот объем информации, который может быть извлечен из фрагмента текста с помощью наших знаний о языке и о мире. В отличие от общей теории аргументации²⁵, в которой исследуются способы доказательства в условиях дискурса и с учетом специфики аудитории, цель прикладного аспекта понятийно-риторического синтеза содержания текстов состоит в выделении частей ценностного композитного высказывания и установлении когерентности тематических фрагментов, позволяющих сделать вывод о коммуникативных задачах и целевой направленности текста.

Композитное высказывание синтезируется путем выделения тематического ядра высказывания, аргументов к нему и умозаключений, предлагаемых говорящим. Получившееся в результате смысловое целое и есть функционально-прагматическое содержание высказывания. Оно складывается из коммуникативных задач высказывания и его целевой направленности.

Коммуникативные задачи высказывания заложены в исходной посылке, т. е. в утверждении, суждении автора, истинность или обо-

²⁵ Ср. [Рузавин 1997]. Риторический подход к исследованию аргументации появился в результате накопления практических знаний о дискурсе, которые невозможно описать средствами формальной логики. В современных работах по основам риторики и теории аргументации отечественные и зарубежные ученые развивают неориторический подход в русле идей, изложенных в работах Хайма Перельмана, см., в частности, [Перельман, Ольбрехт-Тытека 1987].

снованность которого он доказывает в аргументирующей части высказывания. Содержание коммуникативной задачи высказывания соответствует ответам на вопросы: «какую мысль обосновывает автор?», «как автор обосновывает свою мысль?», «как автор характеризует внеязыковую ситуацию?», «какое мнение автор представляет как правильное?», «какие оценки дает автор предмету речи?», «в чем автор пытается убедить аудиторию?», «к чему призывает автор?». Коммуникативная задача соотносится с общей темой высказывания и состоит в ясном изложении какой-либо точки зрения, обосновании какого-либо мнения, в понятном выражении авторской оценки кого- или чего-либо.

Коммуникативные задачи решаются путем установления когерентных связей между структурными элементами высказывания. В структуре композитного высказывания можно выделить несколько элементов, которые при условии когерентности понятийных признаков образуют смысловые взаимосвязи:

- введение темы;
- обозначение проблемы (ситуации);
- обоснование актуальности проблемы;
- характеристика субъекта и объекта проблемы или участников ситуации;
- аргументация важности решения проблемы;
- прогнозирование развития ситуации для субъекта и объекта;
- предложения по решению проблемы.

При определении задач композитного высказывания необходимо выделить исходную мысль, содержащую актуальную для говорящего проблему. В риторической конструкции исходная мысль (посылка) занимает автономную, наиболее независимую позицию, к которой примыкают обосновывающие ее доводы и вывод. Вывод является самым зависимым структурным компонентом композитного высказывания.

Компоненты композитного высказывания могут характеризоваться как по способу речевого воплощения информации о реальности, так и по особенностям функционального использования языковых средств, а именно:

- по особенностям формирования экстенционалов объектов и отражения структуры денотативной ситуации внеязыковой действительности (факты, фактоиды);

- по характеру представления сигнификативной структуры восприятия внеязыковой действительности (мнения);
- по особенностям выражения авторского отношения к внеязыковой действительности (оценки, волеизъявления).

В целом коммуникативные задачи композитного высказывания сводятся к обоснованию, пояснению какой-либо мысли, одобрению, осуждению, оправданию каких-либо состоявшихся или планируемых действий, событий. Например, в тексте настоящей монографии реализуется коммуникативная задача обоснования способов и приемов лингвистического анализа текста.

Целевая направленность композитного высказывания шире, чем его задачи. Она содержится в ответе на вопросы: «для чего обнародован этот текст?», «что хотел сказать автор?», «какой вывод из сказанного предлагает сделать автор?». Свои коммуникативные цели автор воплощает в таких риторических конструкциях, которые интегрируют речевые задания каждого отдельного предложения в интенциональную связность всего высказывания. Наиболее полно целевая направленность высказывания раскрывается в тематическом соотнесении и понятийно-риторической связи авторских выводов с исходными суждениями высказывания и доводами, призванными убедить адресата в правильности и приемлемости утверждений, формирующих, уточняющих или поддерживающих его мнение относительно свойств и качеств объекта высказывания. Факт наличия в содержании высказывания такой соотнесенности вскрывается путем неоднократного установления понятийного единства объекта риторического высказывания и объекта характеристик, которые даются в высказывании. В свою очередь, понятийно-риторическая связь коммуникативных задач и целевой направленности свидетельствует об интенциональных установках высказывания.

Интенциональные установки речевой деятельности можно свести к двум типам: информационному (сообщить факты, передать знания о чем-либо, научить каким-либо действиям, предупредить об ожидаемых или возможных событиях) и регулятивному (оказать управляющее воздействие на поведение людей, склонить аудиторию принять какую-либо точку зрения, признать обоснованным высказанное мнение, побудить адресата предпринять какие-либо действия). Реализованная в тексте интенциональность и есть целевая установка высказывания.

В зависимости от жанровых особенностей текста целевая направленность может быть выражена по-разному. В текстах с преобладанием установки на сообщение фактических сведений качественно-характеризующий и экспрессивно-оценочный компоненты содержания высказывания практически отсутствуют или используются как вспомогательные средства. Например, в тексте «Руководства по изготовлению и применению философского камня» (Парацельс), содержащем определенные фактические сведения и инструкции, реализуется задача последовательного и полного описания необходимых действий, а целевая установка текста — обучение правильному использованию и безотказному применению этого изделия (в том числе и с возможными оздоровительными целями). При этом исследование побудительных мотивов, причин и целей, по которым автор написал, а читатель прочитал такое руководство, не могут быть задачами автоматического анализа текста.

В текстах литературно-художественных и публицистических жанров композитные высказывания с качественно-характеризующим и экспрессивно-оценочным компонентами содержания используются для убеждающего воздействия на аудиторию. Вопрос о том, искренен ли автор в выражении эмоций или они имитируются, для результатов анализа текста представляется несущественным. Важным является то, что качественные, экспрессивные оценки используются авторами как действенный прием усиления, укрепления или изменения взглядов аудитории. Эмоционально-аффективная провокация аудитории является одной из реализаций регулятивной установки автора текста, подтверждающих коммуникативные задачи и целевую направленность текста. Поэтому лингвист, установив наличие эмоциональных коннотаций в тексте, должен указать ту часть диктума, которую говорящий сопровождает дополнительной прагматической информацией. Именно модусные компоненты языковых средств, выражающие личное отношение автора к теме сообщения, позволяют лингвисту делать выводы относительно задач и целей, выраженных в содержании высказывания.

СИНТЕЗ ПОНЯТИЙНО-РИТОРИЧЕСКОЙ СТРУКТУРЫ ТЕКСТА

На этапе синтеза понятийных связей и отношений между позициями по их когерентным семантическим признакам вместе с укрупнением объекта исследования укрупняются единицы иссле-

дования, меняется инструментарий понятийного аппарата. В процедурах синтеза реализуется прагматический подход к интерпретации текста. Функционально-прагматический подход ориентирован на то, чтобы в процессах интерпретации содержания использовать логические и ценностные характеристики текста, аналогичные тем, которые активизированы в сознании слушателя при переработке информации, и фиксировать в явном виде понятийную структуру языковой информации.

Между суждениями, составляющими композитное высказывание, могут устанавливаться два вида связей: подчинительно-обуславливающие (целевые, причинные) и композиционно-симметричные (равноправные, перечислительные) связи. При обуславливающей связи суждений обычным является использование местоимений, синонимов, антонимов, лексических повторов. При параллельной композиционной связи, используемой для описания чего-либо или последовательного повествования о чем-либо, выделяется главный объект высказывания, который характеризуется соположенными (объектными или предикатными) семантическими признаками, связанными с другими объектами отношениями (объединения, перечисления, разделения, противопоставления). Формальная связанность текстовых фрагментов, характеризующаяся единством грамматических форм сказуемого и подтверждающаяся как наличием общей темы, так и единством объекта высказывания, позволяет установить речевые значения слов (контекстное употребление слова).

В качестве отдельного вида связи предложений рассматривают присоединение дополнительного авторского комментария — присоединительную связь (конкатенацию). В отличие от других типов связи, конкатенация не образует самостоятельного текста — это всегда авторское отношение к излагаемому сообщению. Этот вид связи обычно используется в текстах художественной литературы, как средство имитации живой непринужденной речи. В этом случае к последовательной или параллельной связи предложений присоединяется авторская ремарка, выражающая отношение говорящего к сказанному. Например: *Вчера обсуждали их работу. Им крепко досталось — и за дело.*

В отличие от структурной схемы предложения, организующим ядром которой выступает предикат с его актантными связями, в структуре высказывания понятийно-смысловым центром выступает главный риторический объект, характеризующийся в высказыва-

нии. Для того чтобы проанализировать содержательные структуры, характеризующие главный объект речи, необходимо вскрыть семантические связи плана содержания определенного фрагмента текста. Ведь выбор главного риторического объекта не только формирует тему высказывания, но и влияет на понятийно-риторическую структуру текста.

Понятийно-риторическая структура высказывания — реальный лингвистический объект, относящийся к плану содержания, — не вскрывается одним лишь описанием функционально-грамматических взаимоотношений использованных лексических средств. Под структурой понятийного (т.е. мыслительного) содержания понимаются системно-семантические языковые связи, устанавливаемые между отдельными частями полного высказывания с помощью анализа формально-грамматических, лексико-семантических, функционально-синтаксических значений языковых средств воплощения мысли.

Суждения, составляющие композитное высказывание, различаются в зависимости от способа воплощения коммуникативных намерений автора и использованных риторических приемов. Между линейно расположенными суждениями устанавливаются тематические отношения, мотивированные значением обусловленности предикативных признаков пропозиций (причинные, целевые, функциональные отношения) и/или обусловленные когерентными значениями понятийных признаков агентных и объектных актантов, выражающих мереологические, пространственно-временные, перечислительные, сопоставительные отношения (см., например, [Сусов 2006а, 2006б]). И хотя состав и структура суждений, использованных в риторических приемах рассуждения, несколько отличается от высказываний с коммуникативным намерением описания или повествования, тем не менее все композитные высказывания представляют собой трехчастные конструкции.

Исходная посылка, в которой формулируется тема высказывания, обычно представляет собой простое, полное, двусоставное предложение. Такая синтаксическая структура привычна для слушателей, она позволяет легко понять и запомнить содержание сказанного. Исходное положение должно соотноситься с некоторой информационной проблемой (объектом речи), которая обсуждается и обосновывается в последующем рассуждении. Структура суждений, являющихся доводами, может быть распространенной и довольно сложной. Однако их понятийная связь с исходным поло-

жением должна быть понятна слушателям. В качестве обоснования исходного суждения автор может приводить:

- факты, истинность которых подтверждает верность исходной посылки;
- фактоиды, правдоподобие которых подтверждается содержанием посылки;
- оценочные мнения, правильность которых подтверждается совместимостью с прошлым опытом или предпочтительностью с точки зрения ценностей;
- оценки, справедливость которых устанавливается соответствием той или иной норме, соответствием целям и интересам аудитории.

Цель речевого действия заключается в выводе или обобщении высказывания, обычно она формулируется так же кратко, как и посылка.

Если для функционально-понятийного связывания фрагмента текста в качестве исходных единиц синтеза используются пропозиции (факты, мнения, оценки, волеизъявления), то для синтеза понятийно-риторической структуры текста единицами описания являются коммуникативные типы утверждений и суждений: констатирующий, убеждающий, директивный тип. Они могут включать в себя одну или несколько пропозитивных единиц, вскрытых при анализе, а при синтезе коммуникативных типов образуются пропагандистские и агитационные риторические структуры, создающие полное композитное высказывание.

В заключение рассмотрим понятийно-риторическую структуру композитного высказывания убеждающего типа. В полном виде она состоит из трех частей:

- доказываемое положение (исходное суждение) — в форме мнения или факта, являющегося информационным поводом, темой всего высказывания;
- обосновывающий довод — в форме оценочного суждения или оценочного мнения о факте исходной посылки;
- выводное следствие — в форме умозаключения (решения), содержание которого тематически связано с содержанием исходной посылки и обосновывающих доводов.

Следующий пример обладает полной структурой именно такого вида, ср.:

Кабарда испокон веков занимала доминирующее положение в крае. Она не только стояла на более высоком уровне социально-экономического, политического и культурного развития, чем многие народы региона, но она обладала самыми плодородными землями и огромной территорией. Все соседние народы — балкарцы, карачаевцы, осетины, ингуши, чеченцы — занимали в основном горные части Северного Кавказа и всегда ощущали острую нехватку земельных угодий... И только Кабарда могла отводить им эти земли для их поселения... Таким образом, постепенно эти народы оказались в вассальной зависимости от Кабарды²⁶.

Здесь автор дает оценку уровню развития различных этносов, проживавших на территории Северного Кавказа. При этом культурное развитие этнических групп определяется в противопоставлении «более высокий уровень развития — менее высокий уровень развития». Добавим, что этнические группы противопоставляются и по признаку «свой — чужой» с акцентом на исходное суждение — доминирующее положение Кабарды, которая является для автора «своей».

Коммуникативная задача данного композитного высказывания состоит в изложении авторской позиции о том, «кто от кого зависел на Северном Кавказе». Цель высказывания сводится к доказательству исконных прав Кабарды на владение самыми плодородными земельными угодьями Северного Кавказа. (Учитывая этническое многообразие народов Кавказа, тема высказывания и его цель являются конфликтогенными.)

Впрочем, убеждающее высказывание может быть неполным — состоять лишь из двух частей:

- внедряемый в сознание слушателей фактоид — в форме суждения о действиях людей или их личных качествах;
- оценка этого фактоида — в виде оценочной характеристики результатов этих действий или деятельных проявлений личных качеств этих людей.

Впрочем, убеждающее высказывание может быть неполным. В таком случае оно состоит лишь из двух частей: внедряемого в сознание слушателей фактоида (в форме суждения о действиях людей и их личных качествах) и оценки этого фактоида (в виде оценочной характеристики результатов этих действий или деятельных проявлений личных качеств этих людей).

²⁶ Из книги К.Х. Унежева «История Кабарды и Балкарии». Нальчик, 2005. С. 181–182.

ЗАКЛЮЧЕНИЕ

В монографии рассмотрены вопросы автоматического анализа текста в последовательности: от графематического до функционально-прагматического. Кроме формального и смыслового анализа текста, в книге излагается идея последовательного синтеза понятийно-риторических единиц содержания текста и информационного представления текста в метаязыковую структуру данных.

В монографии предлагается оригинальная концепция коммуникативно-синтаксического анализа, которая обусловлена не его объектом, а направленностью на анализ и вскрытие содержания высказывания через распознавание значений слов, составляющих высказывание.

В качестве информационных материалов для работы процессора используются тексты различных прозаических жанров и стилей. В этом отношении предлагаемый процессор является универсальным. При необходимости результаты работы процессора должны позволять дифференциацию анализируемых текстов по признаку жанрово-стилистической принадлежности.

Продукцией коммуникативно-синтаксического анализа служит последовательный список пропозиций (в том числе и свернутых пропозиций), содержащих метаязыковые образы значений слов, занимающих в понятийной структуре высказывания позиции «группы подлежащего» и «группы сказуемого» с выделенными «группой дополнения» и «группой обстоятельств места и времени». Правильное выделение таких групп делает возможным на следующем шаге анализа выявить смысловые связи между предложениями, определить коммуникативные задачи фрагмента текста и его целевую направленность.

В работе используется понятие функционально-смысловых типов речи, дается полная характеристика функциональных и поня-

тийных свойств и особенностей единиц, из которых состоят эти речевые типы.

В книге исследуется проблема временной характеристики протекания действия. Однако наибольшее внимание уделяется процедурам анализа информационной структуры текста. Здесь дается характеристика типов речевой интерпретации действительности, подробно описываются виды анализа текста и последовательность, с которой они используются при работе нашего процессора.

Оригинальной особенностью предложенного подхода к проектированию процессора является разработка и теоретическое обоснование процедур синтеза понятийно-риторических единиц содержания текста.

ИСТОЧНИКИ

- Аверченко А.* Юмористические рассказы. М., 1964.
- Бунин И. А.* Собрание сочинений: в 9 т. М., 1965–1967.
- Вахтин Н. Б., Головкин Е. В.* Социоллингвистика и социология языка: учеб. пособие. СПб., 2004.
- Вересаев В. В.* Пушкин в жизни: в 2 т. М., 2001.
- Вопросы языкознания. 2013. № 5.
- Газовая варочная панель AEG: Инструкция по эксплуатации.
- Газовый духовой шкаф EOG 23400: Руководство по эксплуатации.
- Греция: путеводитель. М., 2011. (Серия «Вокруг света»).
- Даниленко М. П.* Как приготовить дома кондитерские и другие изделия из муки, сладкие блюда, варенье, соки и припасы на зиму (пособие для молодых хозяек). М., 1961.
- Достоевский Ф. М.* Полное собрание сочинений: в 30 т. Л., 1984.
- Ильф И., Петров Е.* Собрание сочинений: в 5 т. М., 1961.
- Казаков Ю. П.* Двое в декабре: Рассказы М., 1966.
- Книга о вкусной и здоровой пище. М., 1952.
- Лермонтов М. Ю. Собрание сочинений: в 4 т. М., 1957–1958.
- Паустовский К.* Собрание сочинений: в 6 т. М., 1957–1958.
- Планеты и созвездия: иллюстрированный справочник. Вильнюс, 2013.
- Польша: путеводитель. М., 2013. (Серия ORANGEвый гид).
- Прага: путеводитель. СПб., 1996. (Серия Polyglott).
- Пушкин А. С.* Полное собрание сочинений: в 10 т. Л., 1977–1979.
- Родос: путеводитель. СПб., 2010. (Серия Nelles Verlag).
- Тургенев И. С.* Три повести: «Ася», «Первая любовь», «Вешние воды». Л., 1947.
- Тургенев И. С.* Записки охотника. М., 1978.
- Чехов А. П.* Собрание сочинений: в 12 т. М., 1960–1964.
- Федеральный закон от 07.02.2011 № 3-ФЗ «О полиции»
- Хармс Д.* Собрание сочинений: в 3 т. М., 2000.
- Respectus Philologicus. 2010. № 17.
- Газета «Метро».

Петербургская интернет-газета «Фонтанка.ру».

<http://news-w.com/62851-18021/n/62851-8631/>

<http://www.24rus.ru>

<http://www.gazeta.spb.ru>

<http://www.rosbalt.ru/>

<http://rusnovosti.ru/news/255776/>

http://www.soprotivlenie.su/slovo/ya_russkiy_kakoy_vostorg.html

<http://slavyanskiysoyuz.ru/26-my-russkie-kakoy-vostorg.html>

[http://gidepark.ru/user/875461648/content/653617\\$](http://gidepark.ru/user/875461648/content/653617$)

<http://rrusia.livejournal.com/5980.html>

<http://www.newsru.com/cinema/29jan2013/filin.html>

<http://bre.ru/news/22763.html>

<http://seltso-vestnik.ru/society/public-associations/2017/11/17/aleksandr-bogomaz-davajte-rabotat-vo-imya-nashej-oblasti-vo-imya-uspeha/>

СЛОВАРИ И СПРАВОЧНИКИ

Баранов О. С. Идеографический словарь русского языка. М., 1995.

БТСРГ = Большой толковый словарь русских глаголов: Идеографическое описание. Синонимы. Антонимы. Английские эквиваленты / под ред. Л. Г. Бабенко. М., 2007.

БТСРС = Большой толковый словарь русских существительных: Идеографическое описание. Синонимы. Антонимы / под ред. Л. Г. Бабенко. М., 2008.

БТС = Большой толковый словарь русского языка / гл. ред. С. А. Кузнецов. СПб., 2009.

Ефремова Т. Ф. Новый словарь русского языка: Толково-словообразовательный. М., 2000.

Золотова Г. А. Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса. М., 1988.

Кузнецова А. И., Ефремова Т. Ф. Словарь морфем русского языка. М., 1986.

Русская грамматика: в 2 т. / гл. ред. Н. Ю. Шведова. М., 1980.

Русский семантический словарь / под ред. Н. Ю. Шведовой. 1998–2007. Т. 1–4.

СРЯ = Словарь русского языка: в 4 т. / под ред. А. П. Евгеньевой. 2-е изд., испр. и доп. М., 1981–1984.

Соловьев Н. В. Русское правописание. Орфографический справочник. 2-е изд., испр. и доп. СПб., 1997.

Стилистический энциклопедический словарь русского языка / под ред. М. Н. Кожинной. 2-е изд., испр. и доп. М., 2006.

Тихонов А. Н. Словообразовательный словарь русского языка: в 2 т. М., 1985.

Hallig R., Wartburg W. von. Begriffssystem als Grundlage für die Lexikographie: Versuch eines Ordnungsschemas. 2 Aufl. Berlin, 1963.

ЛИТЕРАТУРА

- Алексеев П. М. Методика количественной типологии текста. Л., 1983.
- Анисимова Т. В. Типология жанров деловой речи (риторический аспект): автореф. дис. ... д-ра филол. наук. Краснодар, 2000.
- Апресян Ю. Д. Экспериментальное исследование семантики русского глагола. М., 1967.
- Апресян Ю. Д. Нетривиальные семантические признаки и правила выбора значений // Восприятие языкового значения: межвуз. сб. Калининград, 1980.
- Апресян Ю. Д. Избранные труды: в 2 т. Т. 1. Лексическая семантика: 2-е изд., испр. и доп. М., 1995а.
- Апресян Ю. Д. Избранные труды: в 2 т. Т. 2. Интегральное описание языка и системная лексикография. М., 1995б.
- Апресян Ю. Д. Основания системной лексикографии // Языковая картина мира и системная лексикография. М., 2006. С. 32–161.
- Апресян Ю. Д., Богуславский И. М., Иомдин Л. Л., Лазурский А. В., Митюшин Л. Г., Санников В. З., Цинман Л. Л. Лингвистический процессор для сложных информационных систем. М., 1992.
- Апресян Ю. Д., Цинман Л. Л. Перифразирование на компьютере // Семиотика и информатика. Вып. 36. М., 1998. С. 177–202.
- Арефьева С. А. Функционирование языковых средств в официально-деловом стиле // Функционирование языковых единиц и грамматических категорий в разных типах и стилях речи: в 2 ч. Ч. 2. Уфа, 1997. С. 105–110.
- Арутюнова Н. Д. Предложение и его смысл. Логико-семантические проблемы. М., 1976.
- Арутюнова Н. Д. К проблеме функциональных типов лексического значения // Аспекты семантических исследований. М., 1980. С. 156–250.
- Арутюнова Н. Д. Типы языковых значений. Оценка. Событие. Факт. М., 1988.
- Балли Ш. Общая лингвистика и вопросы французского языка. М., 1955.
- Баранов А. Н., Паришин П. Б. Языковые механизмы вариативной интерпретации действительности как средство воздействия на сознание // Роль языка в средствах массовой информации. М., 1986. С. 100–143.
- Бахтин М. М. Вопросы литературы и эстетики. М., 1975.
- Бахтин М. М. Проблема речевых жанров // М. М. Бахтин. Эстетика словесного творчества. М., 1986. С. 250–296.
- Безвербный И. Г. Подходы к решению проблемы определения границ предложения в английском языке // Труды международного семинара «Диалог-2000» по компьютерной лингвистике и ее приложениям. URL: <http://dialog-21.ru/digest/archive/2000/?year=2000&vol=22725&id=6418> (дата обращения: 04.04.2018).
- Белошапкова В. А., Брызгунова Е. А., Земская Е. А. и др. Современный русский язык. 2-е изд., испр. и доп. / под ред. В. А. Белошапковой. М., 1989.

- Бенвенист Э. Общая лингвистика. М., 1974.
- Бердичевский А. С., Иомдин Б. Л. Роль пунктуации в разрешении неоднозначности // Компьютерная лингвистика и интеллектуальные технологии — 2007. URL: <http://www.dialog-21.ru/digests/dialog2007/materials/html/07.htm> (дата обращения: 11.03.2018).
- Бирюлин Л. А. Квантификация предикатов // Типология итеративных конструкций. Л., 1989. С. 249–252.
- Богданов В. В. Семантико-синтаксическая организация предложения. Л., 1977.
- Богданов В. В. Фазисность и фазисные конструкции // Типология конструкций с предикатными актантами. Л., 1985. С. 143–146.
- Богданов В. В. Текст и текстовое общение: учеб. пособие. СПб., 1993.
- Богуславский И. М. Сфера действия лексических единиц. М., 1996.
- Большаков И. А., Большакова Е. И. Автоматический морфоклассификатор русских именных групп // Компьютерная лингвистика и интеллектуальные технологии — 2012. URL: <http://www.dialog-21.ru/digests/dialog2012/materials/pdf/81.pdf> (дата обращения: 11.04.2018).
- Бондарко А. В. Вид и время русского глагола. М., 1971а.
- Бондарко А. В. Грамматическая категория и контекст. Л., 1971б.
- Бондарко А. В. Лимитативность как функционально-семантическое поле // Теория функциональной грамматики. Введение. Аспектуальность. Временная локализованность. Таксис. Л., 1987. С. 45–63.
- Бондарко А. В. Субъектно-предикатно-объектные ситуации // Теория функциональной грамматики. Субъектность. Объектность. Коммуникативная перспектива высказывания. Определенность / неопределенность. СПб., 1992. С. 29–71.
- Бондарко А. В., Буланин Л. Л. Русский глагол. Л., 1967.
- Борисова Е. Г. Дискурсивные слова и референция в процессе понимания сообщения // Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог-2014» (Бекасово, 4–8 июня 2014 г.). URL: <http://www.dialog-21.ru/digests/dialog2014/materials/pdf/BorisovaEG.pdf> (дата обращения: 11.04.2018).
- Брандес М. П. Стилистика немецкого языка. М., 1990.
- Булаховский Л. А. Курс русского литературного языка: в 2 т. Т. 1. Киев, 1952.
- Булыгина Т. В. К построению типологии предикатов в русском языке // Семантические типы предикатов. М., 1982. С. 7–85.
- Булыгина Т. В., Шмелев А. Д. Языковая концептуализация мира. М., 1997.
- Вагурска Б. Структурно-смысловая организация научной статьи. автореф. дис. ... канд. филол. наук. М., 1992.
- Валгина Н. С. Теория текста: учеб. пособие. М., 2003.
- Ван Дейк Т. А. Анализ новостей как дискурса // Ван Дейк Т. А. Язык. Познание. Коммуникация. М., 1989.

- Ванников Ю. В. Типы научных и технических текстов и их лингвистические особенности: метод. пособие: в 2 ч. Ч. 1. М., 1984.
- Васильева А. Н. Газетно-публицистический стиль речи. М., 1982.
- Вежбицка А. Метатекст в тексте // Новое в зарубежной лингвистике. Вып. 8. М., 1978. С. 402–421.
- Вейзе А. А. Чтение, реферирование и аннотирование иностранного текста. М., 1985.
- Веселовская Т. М. Особенности синтаксиса официально-делового стиля // Типология текста в функционально-стилистическом аспекте. Пермь, 1990. С. 119–133.
- Виноградов В. В. Русский язык. Грамматическое учение о слове. М.; Л., 1947.
- Волынец Т. Н. Современный русский язык. Категория рода имени существительного (КСР для студентов филол. фак. специальности D 21.05.02 «Русская филология»). Минск, 2003. URL: <http://www.bsu.by/Cache/pdf/197613.pdf> (дата обращения: 04.04.2018).
- Воробьева М. Б. Некоторые особенности научного произведения обзорного характера // Общие и частные проблемы функциональных стилей. М., 1986. С. 98–117.
- Всеголодова М. В. Теория функционально-коммуникативного синтаксиса. М., 2000.
- Всеголодова М. В., Владимирский Е. Ю. Способы выражения пространственных отношений в современном русском языке. 3-е изд. М., 2009.
- Гак В. Г. К проблеме семантической синтагматики // Проблемы структурной лингвистики. 1971. М., 1972. С. 367–395.
- Гак В. Г. Теоретическая грамматика французского языка. Синтаксис. М., 1986.
- Гальперин И. Р. К проблеме дифференциации стилей речи // Проблемы современной филологии. М., 1965. С. 68–73.
- Гальперин И. Р. Текст как объект лингвистического исследования. М., 1981. Грамматика Пор-Рояля. Л., 1991.
- Грязнухина Т. А. Анализ предложных связей в научном тексте. Киев, 1985.
- Гуртуева Э. М. Деловое письмо. Ташкент, 1991.
- Добров А. В. Автоматическая рубрикация новостных сообщений средствами синтаксической семантики: автореф. дис. ... канд. филол. наук. СПб., 2014.
- Драгой О. В. Влияние объема оперативной памяти на интерпретацию предложения с тремя возможными вершинами придаточного // Компьютерная лингвистика и интеллектуальные технологии — 2006. URL: <http://www.dialog-21.ru/digests/dialog2006/materials/html/Dragoy.htm> (дата обращения: 11.04.2018).
- Дунаевская О. В. О стилях русского языка: учеб. пособие. М., 2010.
- Дюндик Л. Г., Ковалева Л. М., Овсянников Г. И. Неотглагольные предикатные актанты // Типология конструкций с предикатными актантами. Л., 1985. С. 71–72.

- Ермаков А. Е., Пleshко В. В. Компьютерная морфология в контексте анализа связного текста // Материалы международной конференции «Диалог-2004». URL: <http://www.dialog-21.ru/Archive/2004/Ermakov.htm> (дата обращения: 11.04.2018).
- Жанры речи. Вып. 1–3. Саратов, 1997, 1999, 2002.
- Закономерности структурной организации научно-реферативного текста / отв. ред. В. И. Перебейнос. Киев, 1982.
- Зализняк Анна А. Контролируемость ситуации в языке и в жизни // Логический анализ языка. Модели действия. М., 1992. С. 138–145.
- Зализняк Анна А. Семантика кавычек // Компьютерная лингвистика и интеллектуальные технологии — 2007. URL: <http://dialog-21.ru/digests/dialog2007/materials/html/29.htm> (дата обращения: 11.03.2018).
- Золотова Г. А. Коммуникативные аспекты русского синтаксиса. М., 1982.
- Золотова Г. А., Онипенко Н. К., Сидорова М. Ю. Коммуникативная грамматика русского языка. М., 2004.
- Иомдин Б. Л. Многозначные слова в контексте и вне контекста // Вопросы языкознания. 2014. № 4. С. 87–103.
- Иомдин Л. Л. Автоматическая обработка текста на естественном языке: модель согласования. М., 1990.
- Камынина А. А. Современный русский язык. Синтаксис простого предложения. М., 1983.
- Кандрашина Е. Ю., Литвинцева Л. В., Поспелов Д. А. Представление знаний о времени и пространстве в интеллектуальных системах. М., 1989.
- Карпилович Т. П. Семантико-прагматическая структура медиатекста и возможности ее формализации // Контрастивные исследования: текст, предложение, слово. Вып. 2. Минск, 2013. С. 24–31.
- Клейменкова Е. П. Композиционные и стилистические характеристики научной статьи (на материале текстов по экономике): автореф. дис. ... канд. филол. наук. М., 1989.
- Клементьева А. В., Скребцова Т. Г., Суворов С. Г. Графематический анализ текста (опыт разработки) // Материалы XLII Международной филологической конференции (СПбГУ, 11–16 марта 2013 г.). Секция прикладной и математической лингвистики. СПб., 2013. С. 53–60.
- Клышинский Э. С. Некоторые сложности автоматизированной лемматизации несловарных словоформ // Компьютерная лингвистика и интеллектуальные технологии — 2007. URL: <http://www.dialog-21.ru/digests/dialog2007/materials/html/26.htm> (дата обращения: 14.04.2018).
- Кнорина Л. В. Нарушения сочетаемости и разновидности тропов в генитивной конструкции // Логический анализ языка: Противоречивость и аномальность текста. М., 1990. С. 115–124.
- Кобзарева Т. Ю. Морфанализ in vivo // Материалы международной конференции «Диалог-2004». URL: <http://www.dialog-21.ru/Archive/2004/Kobzareva.htm> (дата обращения: 11.04.2018).

- Кобзарева Т. Ю. Омонимия и синонимия знаков препинания в русском тексте // Материалы международной конференции «Диалог-2005». URL: <http://dialog-21.ru/Archive/2005/Kobzareva%20TU/Kobzareva%20TU.htm> (дата обращения: 11.04.2018).
- Кожина А. Н., Крылова О. А., Одинцов В. В. Функциональные типы русской речи: учеб. пособие. М., 1982.
- Кожина М. Н. О специфике художественной и научной речи в аспекте функциональной стилистики. Пермь, 1966.
- Кожина М. Н. О речевой системности научного стиля сравнительно с некоторыми другими. Пермь, 1972.
- Кожина М. Н. О диалогичности письменной научной речи. Пермь, 1986.
- Козинский И. Ш. Кореферентные связи инфинитивных оборотов в русском языке // Типология конструкций с предикатными актантами. Л., 1985. С. 112–116.
- Комарова А. И. Функциональная стилистика: научная речь. Язык для специальных целей (LSP). М., 2004.
- Коньков В. И. Речевая структура газетных жанров: учеб. пособие. СПб., 2004.
- Костомаров В. Г. Русский язык на газетной полосе. М., 1971.
- Красильникова Л. В. Жанр научной рецензии: семантика и прагматика. М., 1999.
- Крылов С. А., Старостин С. А. Актуальные задачи морфологического анализа и синтеза в интегрированной информационной среде STARLING // Материалы международной конференции «Диалог-2003». URL: <http://dialog-21.ru/Archive/2003/Krylov.htm> (дата обращения: 11.03.2018).
- Кудасова О. К. Лингвистические особенности рецензии как разновидности научного текста: автореф. дис. ... канд. филол. наук. М., 1983.
- Курилович Е. Деривация лексическая и деривация синтаксическая // Курилович Е. Очерки по лингвистике. М., 1962. С. 57–70.
- Кустова Г. И. Некоторые проблемы анализа действия в терминах контроля // Логический анализ языка. Модели действия. М., 1992. С. 145–150.
- Кыркунова Л. Г. Стилистические особенности текстов протоколов допроса // Филология на рубеже XX–XXI веков. Пермь, 1996. С. 207–208.
- Лариохина Н. М. Вопросы синтаксиса научного стиля речи. М., 1979.
- Ле Ч. Х. Математическое обеспечение методов распознавания образов при обработке текстов на вьетнамском языке: дис. ... канд. физ.-мат. наук. СПб., 2011.
- Леонтьева Н. Н. Автоматическое понимание текстов: системы, модели, ресурсы: учеб. пособие. М., 2006.
- Лингвистические особенности научного текста. М., 1981.
- Лингвистические распределения (элементы количественного анализа текста). Л., 1985.
- Логический анализ языка. Язык и время / отв. ред. Н. Д. Арутюнова, Т. Е. Янко. М., 1997.

- Лысакова И. П. Язык газеты и типология прессы. Социолингвистическое исследование. СПб., 2005.
- Ляпунова В. Г., Мешков О. Д., Терехова Е. В. Композиция и язык научной рецензии. М., 1990.
- Ляшевская О. Н. К проблеме лемматизации несловарных словоформ // Компьютерная лингвистика и интеллектуальные технологии — 2007. URL: <http://dialog-21.ru/digests/dialog2007/materials/html/62.htm> (дата обращения: 11.03.2018).
- Макаров М. Л. Основы теории дискурса. М., 2003.
- Мартыненко Г. Я. Основы стилеметрии. Л., 1988.
- Марусенко М. А. Атрибуция анонимных и псевдонимных литературных произведений методами распознавания образов. Л., 1990.
- Маслов Ю. С. Глагольный вид в современном болгарском литературном языке (значение и употребление) // Вопросы грамматики болгарского литературного языка. М., 1959. С. 157–312.
- Маслов Ю. С. К основаниям сопоставительной аспектологии // Вопросы сопоставительной аспектологии. Л., 1978. С. 4–44.
- Маслов Ю. С. Очерки по аспектологии. Л., 1984.
- Мелиг Х. Р. Семантика предложения и семантика вида в русском языке (к классификации глаголов Зино Вендлера) // Новое в зарубежной лингвистике. Вып. 15. М., 1985. С. 227–249.
- Мельчук И. А. Опыт теории лингвистических моделей «Смысл ↔ Текст». М., 1974.
- Милославский И. Г. Вопросы словообразовательного синтеза. М., 1980.
- Михеев М. Ю. «Жизни мышья беготня» или «тоска тщетности»? (О метафорической конструкции с родительным падежом) // Вопросы языкознания. 2000. № 2. С. 47–70.
- Мищенко Н. М., Калугина В. В. Об автоматизации составления словарей ложной омонимии флексий // Труды международного семинара «Диалог-2002» по компьютерной лингвистике и ее приложениям. URL: <http://dialog-21.ru/digest/archive/2002/?year=2002&vol=22725&id=7624> (дата обращения: 11.04.2018).
- Москальская О. И. Грамматика текста. М., 1981.
- Муравенко Е. В. Задачи по синтаксической омонимии // Лингвистика для всех. Летние лингвистические школы — 2005 и 2006 / ред.-сост. Е. В. Муравенко, О. Ю. Шеманаева. М.: МЦНМО, 2008. С. 159–162.
- Мурзин Л. Н. Текст как предмет лингвистического исследования // Л. Н. Мурзин, А. С. Штерн. Текст и его восприятие. Свердловск, 1991. С. 6–68.
- Научная литература. Язык. Стилль. Жанры. М., 1985.
- Нечаева О. А. Функционально-смысловые типы речи (описание, повествование, рассуждение). Улан-Удэ, 1974.
- Нечаева О. А. Очерки по синтаксической семантике и стилистике функционально-смысловых типов речи. Улан-Удэ, 1999.

- Никитин М. В. Курс лингвистической семантики: учеб. пособие. СПб., 2007.
- Николаева Т. М. Лингвистика текста // Новое в зарубежной лингвистике. Вып. 8. Лингвистика текста. М., 1978. С. 5–39.
- Николаева Т. М. Функции частиц в высказывании. М., 1985.
- Николина Н. А. Категория времени в художественной речи. М., 2004.
- Нистратова С. Л. Значение и функции местоименно-глагольных конструкций в текстах научного содержания // Функционирование языковых единиц и грамматических категорий в разных типах и стилях речи: в 2 ч. Ч. 2. Уфа, 1997. С. 56–59.
- Новиков А. И. Семантика текста и ее формализация. М., 1983.
- Одинцов В. В. Стилистика текста. М., 1980.
- Орлова Н. В. Жанры разговорной речи и их «стилистическая обработка»: к вопросу о соотношении стиля и жанра // Жанры речи. Вып. 1. Саратов, 1997. С. 51–56.
- Особенности стиля научного изложения. М., 1976.
- Откупщикова М. И. Синтаксис связного текста: учеб. пособие. Л., 1982.
- Падучева Е. В. Высказывание и его соотнесенность с действительностью. М., 1985.
- Падучева Е. В. Семантические исследования. Семантика времени и вида в русском языке. Семантика нарратива. М., 1996.
- Падучева Е. В. Динамические модели в семантике лексики. М., 2004.
- Падучева Е. В. Наблюдатель: типология и возможные трактовки // Компьютерная лингвистика и интеллектуальные технологии — 2006. URL: <http://www.dialog-21.ru/digests/dialog2006/materials/html/Paducheva.htm> (дата обращения: 11.04.2018).
- Падучева Е. В. К интерпретации видо-временных форм в нарративном режиме: настоящее историческое // Компьютерная лингвистика и интеллектуальные технологии. Вып. 9 (16). По материалам международной конференции «Диалог-2010». С. 375–381.
- Падучева Е. В. К семантике русского вида: момент наблюдения и дискурсивный контекст // Компьютерная лингвистика и интеллектуальные технологии: по материалам Международной конференции «Диалог-2016» (Москва, 1–4 июня 2016 г.). URL: <http://www.dialog-21.ru/media/3417/paduchevaev.pdf> (дата обращения: 13.04.2018).
- Паташюте М. Ю. Л. Лексико-синтаксические характеристики научного реферативного текста. автореф. дис. ... канд. филол. наук. М., 1977.
- Перельман Х., Олбрехт-Тытека Л. Из книги «Новая риторика: трактат об аргументации» // Язык и моделирование социального взаимодействия. М., 1987. С. 207–264.
- Пеше М. Контент-анализ и теория дискурса // Квадратура смысла: Французская школа анализа дискурса. М., 1999. С. 105–124.
- Пешковский А. М. Русский синтаксис в научном освещении. М., 1956.
- Попов Э. В. Общение с ЭВМ на естественном языке. 2-е изд. М., 2004.

- Поспелов Н. С. Проблема сложного синтаксического целого в современном русском языке // Уч. зап. МГУ. Вып. 137. М., 1948. С. 32–42.
- Пупынин Ю. А. Активность/пассивность во взаимосвязи с другими функционально-семантическими полями // Теория функциональной грамматики. Персональность. Залоговость. СПб., 1991. С. 211–238.
- Пупынин Ю. А. Грамматические категории русского глагола в их системно-парадигматических и функциональных связях // Межкатегориальные связи в грамматике. СПб., 1996. С. 43–61.
- Радзиевская Т. В. Реферативный текст в лингво-прагматическом аспекте // НТИ. Сер. 2. 1986. № 8. С. 1–5.
- Разновидности и жанры научной прозы. М., 1989.
- Распопов И. П. Строение простого предложения в современном русском языке. М., 1970.
- Ровинская М. Точка как проблема // Труды международного семинара «Диалог-2000» по компьютерной лингвистике и ее приложениям. URL: <http://www.dialog-21.ru/digest/archive/2000/?year=2000&vol=22724&id=6346> (дата обращения: 11.03.2018).
- Розенталь Д. Э. Практическая стилистика русского языка. М., 1998.
- Рубаишкин В. Ш. Представление и анализ смысла в интеллектуальных информационных системах. М., 1989.
- Рузавин Г. И. Методологические проблемы аргументации. М., 1997.
- Румянцева А. В. Комплексный анализ способов выражения множеств лиц в современном русском языке: автореф. дис. ... канд. филол. наук. М., 2009.
- Русская грамматика: в 2 т. / гл. ред. Н. Ю. Шведова. М., 1980.
- Санников В. З. Правила сочетаемости членов русских сочинительных конструкций (СК) // *Lingvistica Silesiana*. № 4. Wrocław, etc., 1981. P. 67–78.
- Санников В. З. Русские сочинительные конструкции. Семантика. Прагматика. Синтаксис. М., 1989.
- Седунов А. А. Модель графематического анализа в системе обработки естественного языка // Вестник ВГУ. Серия: Системный анализ и информационные технологии. 2007. № 2. С. 69–77.
- Селиверстова О. Н. Второй вариант классификационной сетки и описание некоторых предикатных типов русского языка / Семантические типы предикатов. М., 1982. С. 86–157.
- Семантика и синтаксис конструкции с предикатными актантами: мат-лы всесоюзной конференции «Типологические методы в синтаксисе разнотипных языков» (Ленинград, 14–16 апреля 1981 г.). Л., 1981.
- Семиотика и информатика: сб. науч. статей. Вып. 36. М., 1998.
- Серио П. Русский язык и советский политический дискурс: анализ номинализаций // Квадратура смысла: Французская школа анализа дискурса. М., 1999. С. 337–383.
- Сильницкий Г. Г. Семантические типы ситуаций и семантические типы глаголов // Проблемы структурной лингвистики. 1972. М., 1973.

- Сильницкий Г.Г. Структура глагольного значения и результатов // Типология результативных конструкций (результатив, статив, пассив, перфект). Л., 1983. С. 373–391.
- Синдеева Т.И. Некоторые особенности композиционно-речевой организации жанра «научная рецензия» // Функциональные стили и преподавание иностранных языков. М., 1982. С. 105–110.
- Скребцова Т.Г. Семантика глаголов физического действия в русском языке: дис. ... канд. филол. наук. СПб., 1996.
- Скребцова Т.Г. Когнитивная лингвистика: курс лекций. СПб., 2011.
- Скребцова Т.Г. Структурный анализ детективного повествования // Структурная и прикладная лингвистика. Вып. 9. СПб., 2012. С. 80–94.
- Скребцова Т.Г. Определение рода сложносоставных существительных в контексте автоматического морфоанализа // Структурная и прикладная лингвистика. Вып. 10. СПб., 2014. С. 106–112.
- Скребцова Т.Г. К проблеме выделения регулярных семантических компонентов в русской глагольной лексике // Структурная и прикладная лингвистика. Вып. 11. СПб., 2015а. С. 283–295.
- Скребцова Т.Г. Морфоанализ сложносоставных существительных в русском языке: проблема определения рода и одушевленности // XLIII Международная филологическая конференция (Санкт-Петербург, 11–16 марта 2014 г.): избранные труды. СПб.: СПбГУ, 2015б. С. 386–395.
- Сокирко А.В. Быстрословарь: предсказание морфологии русских слов с использованием больших лингвистических ресурсов // Компьютерная лингвистика и интеллектуальные технологии — 2010. URL: <http://www.dialog-21.ru/digests/dialog2010/materials/html/68.htm> (дата обращения: 04.04.2018).
- Солганик Г.Я. Синтаксическая стилистика. Сложное синтаксическое целое. М., 1973.
- Солганик Г.Я. Очерки модального синтаксиса. М., 2010.
- Стародумова Е.А. Синтаксис современного русского языка. Владивосток, 2005.
- Статистика речи и автоматический анализ текста. Л., 1973.
- Статистичні параметри стилів. Киев, 1967.
- Сусов А.А. Каузальные риторические отношения // Тверской лингвистический меридиан. Вып. 6. Тверь, 2006а. С. 55–81.
- Сусов А.А. Моделирование дискурса в терминах теории риторической структуры // Вестник Воронеж. гос. ун-та. Серия: Филология. Журналистика. 2006б. № 2. С. 133–138.
- Талми Л. Отношение грамматики к познанию // Вестник Московского университета. Сер. 9. Филология. 1999. № 1, 4, 6.
- Татарникова Н.М. Координация первичного и вторичного речевых жанров в официально-деловом стиле речи. Иркутск, 2006.
- Телин Н.Б. Вид и способ действия в русском языке // Новое в зарубежной лингвистике. Вып. 15. М., 1985. С. 250–260.

- Тестелец Я. Г. Введение в общий синтаксис. М., 2001.
- Типология конструкций с предикатными актантами / отв. ред. В. С. Храковский. Л., 1985.
- Тискин Д. Б. «Аппозиционные» и «соопределяющие» клаузы: к вопросу о локализации условной семантики // Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог-2016» (Москва, 1–4 июня 2016 г.). URL: <http://www.dialog-21.ru/media/3434/tiskindb.pdf> (дата обращения: 13.04.2018).
- Троянская Е. С. Обучение чтению научной литературы. М., 1989.
- Тузов В. А. Компьютерная семантика русского языка. СПб., 2003.
- Тураева З. Я. Категория времени: время грамматическое и время художественное (на материале английского языка): учеб. пособие. 2-е изд. М., 2009.
- Уразлин Ю. К. Анализ слабо структурированных текстовых данных: автореф. дис. ... канд. тех. наук. М., 2005.
- Успенский Б. А. Поэтика композиции. М., 1970.
- Учебно-методический комплекс по курсу «Синтаксис современного русского языка»: препринт учебника для высших учебных заведений Российской Федерации. СПб., 2008.
- Федорова О. В., Янович И. С. Об одном типе синтаксической неоднозначности, или Кто стоял на балконе // Материалы международной конференции «Диалог-2004». URL: <http://www.dialog-21.ru/Archive/2004/Fedorova%20O.htm> (дата обращения: 11.04.2018).
- Филиппов К. А. Лингвистика текста: курс лекций. СПб., 2003.
- Филлмор Ч. Дело о падеже // Новое в зарубежной лингвистике. Вып. 10. М., 1981. С. 369–495.
- Холодович А. А. Категория множества в японском в свете теории множества в языке // А. А. Холодович. Проблемы грамматической теории. Л., 1979. С. 173–195.
- Хорохордина О. В. Инструкция как тип текста // Мир русского слова. 2013. № 4. С. 7–14.
- Храковский В. С. Семантика фазовости и средства ее выражения // Теория функциональной грамматики. Введение. Аспектуальность. Временная локализованность. Таксис. Л., 1987. С. 153–180.
- Храковский В. С. Семантические типы множества ситуаций и их естественная классификация // Типология итеративных конструкций. Л., 1989. С. 5–53.
- Храковский В. С. Грамматические категории глагола (опыт теории взаимодействия) // Межкатегориальные связи в грамматике. СПб., 1996. С. 22–42.
- Цветкова Т. М. Опыт семантико-синтаксического анализа текста типа «рассуждение» // Сборник научно-методических материалов к лекциям для слушателей ФПК. М., 1982. С. 234–256.
- Чебанов С. В., Мартыненко Г. Я. Семиотика описательных текстов. СПб., 1999.
- Чернышков Д. М. Об одном статистическом методе пополнения морфологического словаря // Компьютерная лингвистика и интеллектуальные

- технологии — 2010. URL: <http://dialog-21.ru/digests/dialog2010/materials/html/80.htm> (дата обращения: 04.04.2018).
- Шаляпина З. М. О двух типах условий кореферентности валентностей в предикатно-актантных конструкциях // Типология конструкций с предикатными актантами. Л., 1985. С. 46–50.
- Шварцкопф Б. С. Официально-деловой язык // Культура русской речи и эффективность речевого общения. М., 1996. С. 270–281. С. 131.
- Шведова Н. Ю. Детерминант // Лингвистический энциклопедический словарь / гл. ред. В. Н. Ярцева. М., 1990.
- Шенк Р. Обработка концептуальной информации. М., 1980.
- Шмелев Д. Н. Проблемы семантического анализа лексики (на материале русского языка. М., 1973).
- Шмелев Д. Н. Синтаксическая членимость высказывания в современном русском языке. М., 1976.
- Шмелев Д. Н. Русский язык в его функциональных разновидностях. М., 1977.
- Шмид В. Нарратология. М., 2003.
- Юдина М. В. Понимание и порождение высказываний с синтаксической неоднозначностью (на примере относительных придаточных в русском языке) // Компьютерная лингвистика и интеллектуальные технологии — 2006. URL: <http://www.dialog-21.ru/digests/dialog2006/materials/html/Yudina.htm>
- Язык и стиль научного изложения. М., 1983.
- Язык научной литературы. М., 1975.
- Якобсон Р. В поисках сущности языка // Семиотика. М., 1983. С. 102–117.
- Яковлева Е. С. Фрагменты русской языковой картины мира (модели пространства, времени и восприятия). М., 1994.
- Яцко В. А. Особенности семантико-синтаксической структуры аннотативных высказываний // НТИ. Сер. 2. 1993. № 5. С. 20–22.
- Яцко В. А. Рассуждение как тип научной речи. Абакан, 1998.
- Beaugrande R., de. Text, Discourse, and Process: Toward a Multidisciplinary Science of Texts. Norwood, NJ, 1980.
- Bell A. The discourse structure of news stories // Approaches to Media Discourse. Oxford, 1998. P. 64–104.
- Coulthard M. An Introduction to Discourse Analysis. London, 1977.
- Fauconnier G. Mental Spaces: Aspects of Meaning Construction in Natural Languages. Cambridge, MA, 1985.
- Fillmore Ch. J. Types of lexical information // F. Kiefer (ed.). Studies in Syntax and Semantics. Dordrecht, 1969. P. 109–137.
- Gibbs R. W. What's cognitive about cognitive linguistics // E. H. Casad (ed.). Cognitive Linguistics in the Redwoods: The Expansion of a New Paradigm in Linguistics. Berlin; New York, 1996. P. 27–53.
- Halliday M. A. K., Hasan R. Cohesion in English. London; New York, 1976.

- Labov W., Waletzky J.* Narrative analysis // J. Helm (ed.) Essays on the Verbal and Visual Arts. Seattle, 1967. P. 12–44.
- Lakoff G.* Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. Chicago, 1987.
- Lakoff G., Johnson M.* Metaphors we live by. Chicago, 1980.
- Langacker R. W.* A view of linguistic semantics // B. Rudzka-Ostyn (ed.). Topics in cognitive linguistics. Amsterdam; Philadelphia, 1988. P. 49–90.
- Langacker R. W.* Concept, Image, and Symbol: The Cognitive Basis of Grammar. Berlin; New York, 1991.
- Levinson S. C.* Space in Language and Cognition: Explorations in Cognitive Diversity. Cambridge, 2003.
- Palmer F. R.* Semantics: A New Outline. M., 1982.
- Stubbs M.* Discourse Analysis: The Sociolinguistic Analysis of Natural Language. Chicago, 1983.
- Talmy L.* How language structures space // H. L. Pick, Jr., L. P. Acredolo (eds.). Spatial Orientation: Theory, Research, and Application. New York; London, 1983. P. 225–282.

СПИСОК СОКРАЩЕНИЙ В ПРИМЕРАХ

буд	— будущее время
вин	— винительный падеж
вопр	— вопросительное
глаго	— глагол
дат	— дательный падеж
ед	— единственное число
жр	— женский род
им	— именительный падеж
колич	— количественное
межд	— междометие
мест	— местоимение
мн	— множественное число
мр	— мужской род
нар	— наречие
наст	— настоящее время
неизм	— неизменяемый
неод	— неодушевленное
неперех	— непереходный
несоверш	— несовершенный вид
од	— одушевленное
отч	— отчество
перех	— переходный
превосх	— превосходная степень
пред	— предложный падеж
предл	— предлог
прил	— прилагательное
прич	— причастие
прош	— прошедшее время
род	— родительный падеж
соверш	— совершенный вид
ср	— средний род
сущ	— существительное
тв	— творительный падеж
част	— частица
числ	— числительное
фам	— фамилия

Приложение 1

АНАЛИЗ ПРЕДЛОЖЕНИЙ ИЗ ПОВЕСТИ Ю. В. ТРИФОНОВА «ОБМЕН»

ТЕКСТ

В июле мать Дмитриева Ксения Федоровна тяжело заболела, и ее отвезли в Боткинскую, где она пролежала двенадцать дней с подозрением на самое худшее. В сентябре сделали операцию, худшее подтвердилось, но Ксения Федоровна, считавшая, что у нее язвенная болезнь, почувствовала улучшение, стала вскоре ходить, и в октябре ее отправили домой, пополневшую и твердо уверенную в том, что дело идет на поправку. Вот именно тогда, когда Ксения Федоровна вернулась из больницы, жена Дмитриева затеяла обмен: решила срочно съезжаться со свекровью, жившей одиноко в хорошей, двадцатиметровой комнате на Профсоюзной улице.

Графематический анализ

Графематический анализ данного текста не представляет трудностей, так как в нем нет сокращений, цифровых последовательностей, прямой речи. Все слова составлены из символов одного алфавита и отделены друг от друга пробелами или знаками препинания. Предложения и части предложений также выделяются однозначно.

Морфологический анализ

Все словоформы текста содержатся в словаре словоформ и, следовательно, получают морфологическую атрибуцию, правда, не всегда однозначную (варианты приводятся через косую черту).

в — предл
июле — сущ, мр, пр, ед
мать — сущ, жр, им, ед / сущ, жр, вин, ед / межд
Дмитриева — сущ, фам, жр, им, ед / сущ, фам, мр, вин, ед / сущ, фам, мр, род, ед
Ксения — сущ, имя, мр, род, ед / сущ, имя, мр, вин, ед / сущ, имя, жр, им, ед
Федоровна — сущ, отч, жр, им, ед
тяжело — прил кратк, ср, ед / нар
заболела — глаг, неперех, соверш, прош, жр, ед

и — союз/междом
ее — мест. прил, неизм/мест, 3 л, жр, род, ед/мест, 3 л, жр, вин, ед
отвезли — глаг, перех, соверш, прош, мн
в — предл
Боткинскую — прил, жр, вин, ед
где — нар, вопр/союз/част
она — мест, 3 л, жр, им, ед
пролежала — глаг, неперех, соверш, прош, жр, ед/глаг, перех, соверш, прош, жр, ед
двенадцать — числ, колич, им/числ, колич, вин
дней — сущ, мр, род, мн
с — предл
подозрением — сущ, ср, тв, ед
на — предл/част
самое — мест. прил, неизм/мест. прил, ср, им, ед/мест. прил, ср, вин, ед
худшее — сущ, ср, вин, ед/сущ, ср, пр, ед/сущ, ср, им, ед/прил, превосх, ср, вин, ед/прил, превосх, ср, им, ед

Конструктивно-синтаксический анализ

Перед тем как начать синтаксический анализ, лингвистический процессор обращается к Семантико-синтаксическому словарю, откуда (по лемме) получает соответствующие характеристики словоформ.

1. Синтаксические группы, обрабатываемые правилами первого ранга

В июле — однозначно собирается как предложно-падежная группа, потому что предлог *в* имеет единственную грамматическую характеристику. А из семантико-синтаксического словаря мы получаем порядка 30 значений для управления винительным падежом, и порядка 20 значений для управления предложным падежом. Так, как словоформа *июле* имеет только одно грамматическое значение и один семантико-синтаксический вариант 'промежуток времени', то из всех значений предлога *в* нам подойдет только то, в котором он управляет предложным падежом и сочетается со словами темпоральной семантики. Следовательно, данная падежно-предложная группа является обстоятельством времени и относится к предложению в целом.

мать Дмитриева Ксения Федоровна — здесь мы сталкиваемся с неоднозначностью. Из семантико-синтаксического словаря мы получаем несколько значений для слова *мать*, причем часть из них не проходит через фильтр и отсекается. К ним относятся слово *мать* в функции обращения (должно быть выделено запятыми), а также значения, связанные с олицетворением неодушевленных существностей (*Родина-мать*, *мать-природа* и т.п.), поскольку требуется наличие дефиса. Остаются следующие три значения: 1) 'женщина по отношению к своим детям', 2) 'самка по отношению к своим детенышам', 3) 'жена священника или монахиня'. Обозначение человека по

родственным отношениям может сопровождаться фамилией, именем, отчеством в различных комбинациях, которые согласуются по роду, числу, падежу с главным словом. Напротив, это нетипично для третьего и тем более второго значений. Так как форма отчества атрибутируется однозначно, это ограничивает выбор падежа для слов *мать* и *Ксения* — остается только именительный падеж. А вот фамилия в силу своей грамматической многозначности может принимать участие в двух разных семантико-синтаксических конструкциях:

мать (кого) *Дмитриева* (мр, род, ед) + имя и отчество матери

мать + *Дмитриева* (жр, им, ед) = фамилия матери + имя и отчество матери

В обоих вариантах мы имеем дело с аппозитивной конструкцией. В первом случае в качестве определяемого компонента выступает словосочетание *мать Дмитриева*, а аппозитивом — имя и отчество. Во втором случае определяемым словом является *мать*, а аппозитивом — фамилия, имя и отчество. (По правилам полное Ф. И. О. здесь должно было быть выделено запятыми или тире, но на практике это соблюдается не всегда. Поэтому мы вынуждены рассматривать оба варианта.) Различие между двумя указанными вариантами существенно с точки зрения дальнейшего (коммуникативно-синтаксического) анализа, так как при первой интерпретации требуется создавать в информационной структуре два объекта — «Дмитриева» и «его мать», а во втором — только один («мать», причем по результатам анализа первого предложения неизвестно чья). При анализе последующего текста эта неоднозначность будет снята, так как «Дмитриев» станет фигурировать как самостоятельный объект.

тяжело заболела — словоформа *тяжело* перед личной формой глагола может быть только наречием в функции обстоятельства образа действия. В семантико-синтаксическом словаре *заболеть* характеризуется как глагол процесса; впоследствии на основе этой информации будет сделан вывод, что *мать* — это не активный деятель, а экспериенцер такого процесса, как болезнь. Кроме того, глаголу *заболеть* приписывается значение 'начало' по метаязыковому признаку «фазовость». Поскольку глагол *заболеть* не управляет именной группой в творительном падеже («заболеть чем»), можно исключить значение, не связанное со здоровьем (например, «заболеть футболом»). Таким образом, в данном случае речь идет о здоровье, без указания на конкретную болезнь.

ее отвезли — в силу переходности глагола мы получаем, что единственный претендент на прямое дополнение — словоформа *ее*. Тем самым она становится однозначным (личное местоимение в винительном падеже). Также появляется фиктивное подлежащее «какие-то люди».

в Боткинскую — у прилагательного *Боткинская* нет согласованного существительного, значит, данное прилагательное само выступает в функции существительного. Наличие рядом глагола перемещения *отвезли*, так же

как союзного слова *где*, позволяет из многочисленных значений предлога *в* (+ $S_{\text{вин.}}$), сделать заключение, что выражение в *Боткинскую* означает «обстоятельство места-куда».

она пролежала — словоформа *она* становится кандидатом на подлежащее (согласована с личным глаголом). Возникает проблема неоднозначности глагола. Значение, связанное с воздействием на предмет (например, «пролежать диван») исключается. Остаются еще два: «побыть больным, на излечении» и «пробыть какое-то время в лежачем состоянии». Тема болезни подталкивает к выбору первого значения.

двенадцать дней — конструкция собирается по общему правилу: «количественное числительное + родительный падеж исчисляемого объекта». Вариант именительного падежа придется исключить, так как в данном придаточном уже есть словоформа в именительном падеже (*она*).

с подозрением (на самое худшее) — предложно-падежная группа *на самое худшее* стоит в конце предложения, поэтому без вариантов собирается в единый элемент (предлог *на* может управлять винительным падежом, а компоненты словосочетания *самое худшее* согласованы друг с другом по роду, числу, падежу). *Подозрение* — отглагольное существительное, которое в одном из своих значений может иметь при себе дополнение в «на + $S_{\text{вин.}}$ ». Далее, правда, встает вопрос о том, у кого было это подозрение: у Ксении Федоровны или у кого-то другого (врачей?). Но эта информация в тексте отсутствует.

1. Синтаксические группы, обрабатываемые правилами второго и третьего ранга

она пролежала двенадцать дней с подозрением на самое худшее — получаем полное предложение с подлежащим, сказуемым, прямым дополнением и примыкающей предложно-падежной группой. Данную конструкцию мы можем собрать в силу того, что с графематической точки зрения она представляет собой «часть предложения», которая стоит в конце всего предложения (следовательно, за ней не может быть ничего подчиненного).

в *Боткинскую, где* — из возможных частей речи для слова *где* выбираем союз (союзное слово), так как оно стоит после запятой. Далее из возможных вариантов интерпретации союза выбираем союз, присоединяющий придаточное определительное предложение (для придаточного места требуется наличие непосредственно перед запятой указательных слов *там, туда, оттуда*; для придаточного дополнительного — наличие глагола ментальной или речевой деятельности; придаточное уступительное занимает позицию перед главным предложением). Присоединяем собранную группу к определённому слову («обстоятельству места-куда»).

Ее отвезли в Боткинскую — отсутствие предлога *от* после глагола *отвезти* позволяет отбросить интерпретацию «увезти в сторону от чего-л.», так что глагол становится однозначным. В результате собирается неопределен-

но-личное предложение с глаголом направленного объектного перемещения, прямым дополнением *ее* и локативом «куда».

В июле мать... заболела — в силу непереходности глагола мы имеем только один вариант сборки — подлежащее + сказуемое + обстоятельство времени (характерное для фазового глагола).

и — не может быть междометием, так как не выделено запятыми, значит, это союз. Поскольку личная форма глагола после запятой и данного союза (*отвезли*) не совпадает по форме с личной формой глагола до запятой (*заболела*), они не могут быть однородными сказуемыми. Значит, союз *и* соединяет две части сложносочиненного предложения. Вопрос заключается в том, как интерпретировать связь между событиями, описанными в этих частях: как исключительно темпоральную или вдобавок как причинно-следственную.

«предложение, союз + предложение (конец предложения)» — сборка сложного предложения.

Коммуникативно-синтаксический анализ

1) Объекты:

1 — человек женского рода = мать (родственная функция к объекту_2), Ксения Федоровна (собственные имя и отчество), свекровь (родственная функция к объекту_5)

2 — человек мужского рода = Дмитриев (собственная фамилия), (родственная функция к объекту_1), (родственная функция к объекту_5)

3 — организация лечебная = Боткинская (название), больница

4 — худшее (?? диагноз, болезнь)

5 — человек женского рода = жена (родственная функция по отношению к объекту_2), (родственная функция по отношению к объекту_1), Лена (собственное имя)

6 — помещение = комната

7 — локатив = улица, Профсоюзная (название)

8 — люди = кто-то, кто ее отвез

9 — люди = кто-то, кто имел нехорошее предчувствие

NB: Объект_9 может совпадать с объектом_1 или с объектом_8 либо быть отличным от обоих.

2) Признаки:

1 — процесс = болезнь, начало, тяжело — сильно

2 — перемещение = отвезли

3 — процесс = пролежала

4 — состояние = подозрение

5 — ??? ожидание, диагноз = худшее

3) Время:

1 — июль

2 — когда отвезли

3 — двенадцать дней

4) Место:

1 — начальное местоположение

2 — объект_3

5) Пропозиции:

Пропозиция 1: объект_1 имеет признак_1 (процесс) во время_1 [женщина с именем и родственными отношениями + начала, болеть, тяжело + в июле в месте_1]

Пропозиция 2: объект_1 имеет признак_2 в месте_2 (= объект_3) во времени после времени_1 (время_2) [кто-то + переместил + женщину + в Боткинскую + после или из-за пропозиции_1]

Пропозиция 3: объект_1 имеет признак_3, в месте_2 в течение времени_3 (после времени_2) [женщина + пролежала + в Боткинской + две недели (после пропозиции_2)]

Пропозиция 4: объект_9 или объект_1 имеет признак_4 в месте_2 (в течение времени_3) [кто-то или сама женщина + подозревали худшее + в Боткинской + двенадцать дней (в течение пропозиции_3)]

Приложение 2

ПРИМЕРЫ СТАТЕЙ СЕМАНТИКО-СИНТАКСИЧЕСКОГО СЛОВАРЯ

СУЩЕСТВИТЕЛЬНЫЕ

Семантическое описание существительного обычно ограничивается описанием его собственного значения. Прежде всего выделяется базовый семантический признак со значениями 'живое существо', 'предмет', 'действие', 'время', 'количество', 'качество' и некоторыми другими. Также может быть задействован набор дополнительных семантических признаков (см. гл. 3). Выделение признаков происходило эмпирическим путем и на протяжении работы их набор претерпевал изменения. Основным критерием выделения того или иного признака (или значения признака) является его способность дифференцировать разные ССВ одного слова, т. е. эксплицировать те компоненты конкретного значения слова, которые отличают его от других значений того же слова.

Приведем некоторые примеры.

ДОЛГ, = Обязанность перед кем-л., чем-л.

(собств. знач.) сем. признаки: базовый: действие

ДОЛГ, = То, что взято или отдано заимообразно (преимущественно деньги)

(собств. знач.) сем. признаки: базовый: предмет

ВЛАДѢНИЕ, = к Владеть

(собств. знач.) сем. признаки: базовый: действие

ВЛАДѢНИЕ, = Территория, принадлежащая кому-л., находящаяся в чьей-л. собственности

(собств. знач.) сем. признаки: базовый: место

ДА́ЧА, = Загородный дом для летнего отдыха

(собств. знач.) сем. признаки: базовый: предмет

ДА́ЧА, = Загородная местность с таким домом (домами)

(собств. знач.) сем. признаки: базовый: место

ДОБРО́, = Хорошее, доброе дело, поступок

(собств. знач.) сем. признаки: базовый: действие; оценка = полож.

ДОБРО́, = Вещи, имущество, скарб

(собств. знач.) сем. признаки: базовый: предмет

ДОСТОИНСТВО, = Положительное качество

(собств. знач.) сем. признаки: базовый: свойство, оценка = полож.

ДОСТОИНСТВО, = Внешнее проявление самоуважения, сознания своей значимости

(собств. знач.) сем. признаки: базовый: действие

ДОСТОИНСТВО, = Стоимость, ценность (денежных знаков и ценных бумаг)

(собств. знач.) сем. признаки: базовый: количество

ДОСТОИНСТВО, = Титул, чин, звание

(собств. знач.) сем. признаки: базовый: ментально-информационный

ДЕПУТАТ, = Выборный представитель, член выборного государственного учреждения

(собств. знач.) сем. признаки: базовый: живое существо; человек = +; взрослый = +

ДЕПУТАТ, = Лицо, избранное кем-л. и уполномоченное для выполнения каких-л. поручений, заданий и т. п.; делегат

(собств. знач.) сем. признаки: базовый: живое существо; человек = +; взрослый = +; релятивность = +

Разумеется, потенциал семантических признаков не столь высок, чтобы разграничить все значения любого слова. В ряде случаев возможна лишь частичная дифференциация, ср.:

БОЛЕЗНЬ, = Конкретное заболевание, нарушающее деятельность организма (или его отдельных органов); расстройство здоровья

(собств. знач.) сем. признаки: базовый: действие

БОЛЕЗНЬ, = О нарушениях или отклонениях в жизнедеятельности животных, растений

(собств. знач.) сем. признаки: базовый: действие

БОЛЕЗНЬ, = Разрушение, повреждение чего-л.

(собств. знач.) сем. признаки: базовый: действие

БОЛЕЗНЬ, = Отклонение от нормы в чем-л.; отрицательное качество, свойство

(собств. знач.) сем. признаки: базовый: свойство, оценка = отриц.

ВЫПУСК, = к Выпускать и Выпускаться

(собств. знач.) сем. признаки: базовый: действие

ВЫПУСК, = Часть литературного или научного произведения, издаваемая отдельной книгой, брошюрой

(собств. знач.) сем. признаки: базовый: предмет

ВЫПУСК, = Группа учащихся (класс, курс), окончивших учебное заведение одновременно

Приложение 2

(собств. знач.) сем. признаки: базовый: живое существо; человек = +;
взрослый = +

ВЫПУСК, = О партии, серии изделий одной даты изготовления

(собств. знач.) сем. признаки: базовый: предмет

ДЕТИ, = Мальчики и (или) девочки до 14–16 лет

(собств. знач.) сем. признаки: базовый: живое существо; человек = +;
взрослый = –

ДЕТИ, = Сыновья или дочери (независимо от возраста)

(собств. знач.) сем. признаки: базовый: живое существо; человек = +;
взрослый = –; релятивность = +

ДЕТИ, = О наивных, неопытных людях

(собств. знач.) сем. признаки: базовый: живое существо; человек = +;
взрослый = +

ДЕТИ, = О людях, являющихся характерными представителями какой-л.
среды, эпохи и т. п., тесно связанных с кем-, чем-л.

(собств. знач.) сем. признаки: базовый: живое существо; человек = +;
взрослый = +

Следует признать, что не все значения слов удалось дифференцировать
при помощи текущих семантических признаков и их значений, ср.:

АРЕСТ, = Заключение под стражу, лишение свободы

(собств. знач.) сем. признаки: базовый: действие

АРЕСТ, = Запрещение распоряжаться имуществом или денежными сред-
ствами, налагаемое судебными органами

(собств. знач.) сем. признаки: базовый: действие

ЛЕТА, = Годы

(собств. знач.) сем. признаки: базовый: время

ЛЕТА, = Возраст

(собств. знач.) сем. признаки: базовый: время

ГЛАГОЛЫ

В отличие от существительных, в семантическом описании глагола основную роль играют условия реализации, характеризующие возможности заполнения его валентностей (с синтаксической и семантической точек зрения). При описании семантических характеристик глагольных актантов использовался тот же набор семантических признаков и их значений, при помощи которых описываются значения существительных (см. выше).

Приведем некоторые примеры.

ВОЗЛОЖИТЬ, = Торжественно положить сверху, поверх чего-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем.

признаки: базовый: живое существо; человек = +; второй актант: падеж = винительный / предлог = без предлога / сем. признаки: базовый: предмет; третий актант: падеж = винительный / предлог = «на» / сем. признаки: базовый: предмет

ВОЗЛОЖИТЬ, = Поручить кому-л. что-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = винительный / предлог = без предлога / сем. признаки: базовый: абстрактное; третий актант: падеж = винительный / предлог = «на» / сем. признаки: базовый: живое существо; человек = +

НУЖДАТЬСЯ, = Жить в бедности, испытывать нужду (1 зн.).

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +

НУЖДАТЬСЯ, = Испытывать потребность в ком-, чем-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = предложный / предлог = «в» / сем. признаки: базовый: любой

НУЖДАТЬСЯ, = Требовать чего-л., вызывать необходимость чего-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: любой; второй актант: падеж = предложный / предлог = «в» / сем. признаки: базовый: действие

ПРОТИВОРЕЧИТЬ, = Возражать, не соглашаться с кем-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = дательный / предлог = без предлога Косвенное дополнение — дательный падеж без предлога / сем. признаки: базовый: живое существо; человек = +

ПРОТИВОРЕЧИТЬ, = Находиться в противоречии с чем-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: абстрактное; второй актант: падеж = дательный / предлог = без предлога / сем. признаки: абстрактное

ОКРУЖИТЬ, = Расположиться, занять место вокруг кого-, чего-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; второй актант: падеж = винительный / предлог = без предлога / сем. признаки: базовый: живое существо ИЛИ базовый: предмет

ОКРУЖИТЬ, = Поместить, расположить что-л. вокруг кого-, чего-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = винительный / предлог = без предлога / сем. признаки: базовый: предмет ИЛИ базовый: место; третий актант: падеж = творительный / предлог = без предлога / сем. признаки: базовый: предмет

ОКРУЖИТЬ, = Приблизить к кому-л. каких-л. лиц, группу лиц.
(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = винительный / предлог = без предлога / сем. признаки: базовый: живое существо; человек = +; третий актант: падеж = творительный / предлог = без предлога / сем. признаки: базовый: живое существо; человек = +

ОКРУЖИТЬ, = Создать вокруг кого-л. какую-л. обстановку, условия, установить к кому-л. то или иное отношение

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = винительный / предлог = без предлога / сем. признаки: базовый: живое существо; третий актант: падеж = творительный / предлог = без предлога / сем. признаки: базовый: абстрактное

Примеры частичной дифференциации значений при помощи семантических признаков, ср.:

ВЫЕХАТЬ, = Уехать, отправиться откуда-л., куда-л.
(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: обстоятельство_места_откуда_куда¹

ВЫЕХАТЬ, = Оставить прежнее местожительство
(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +;

ВЫЕХАТЬ, = Прибыть куда-л. (на лошади, в экипаже, машине и т. п.); появиться где-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: обстоятельство_места_куда

ВЫЕХАТЬ, = Использовать в своих интересах, извлечь выгоду из кого-, чего-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = предложный / предлог = «на» / сем-признаки: базовый: живое существо; человек = + ИЛИ базовый: абстрактное

ВЫЕХАТЬ, = Вылезти, высунуться, показаться из-под чего-л. (об одежде).
(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: предмет; второй актант: падеж = родительный / предлог = «на» ИЛИ предлог = «из-под» / сем. признаки: базовый: предмет

¹ Общий термин, который применительно к конкретному глаголу соответствует перечню возможных конструкций. См. также ниже.

ОБВИНЯТЬ, = Считать виноватым, виновным в чем-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = винительный / предлог = без предлога / сем. признаки: базовый: живое существо; третий актант: падеж = винительный / предлог = «в» ИЛИ «за» / сем. признаки: базовый: действие

ОБВИНЯТЬ, = Упрекать в чем-л., осуждать за что-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = винительный / предлог = без предлога / сем. признаки: базовый: живое существо; человек = +; третий актант: падеж = предложный / предлог = «в» / сем. признаки: базовый: абстрактное

ОБВИНЯТЬ, = Считая виновным в чем-л., привлекать к судебному разбирательству; судебным порядком устанавливать виновность

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = винительный / предлог = без предлога / сем. признаки: базовый: живое существо; человек = +; третий актант: падеж = предложный / предлог = «в» / сем. признаки: базовый: действие

ОБВИНЯТЬ, = Поддерживать обвинение, выступать в качестве судебного обвинителя, прокурора

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +

Примеры лемм, которые не удалось дифференцировать с помощью выделенных признаков, ср.:

ОПУБЛИКОВАТЬ, = Напечатать для всеобщего сведения, сделать известным путем помещения в печати

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = винительный / предлог = без предлога / сем. признаки: базовый: ментально-информационный; третий актант = обстоятельство_места_где

ОПУБЛИКОВАТЬ, = Печатать массовым тиражом, издавать какое-л. произведение, текст

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = винительный / предлог = без предлога / сем. признаки: базовый: ментально-информационный

СОБЛЮДАТЬ, = Строго придерживаться чего-л., в точности выполнять что-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант:

падеж = винительный / предлог = без предлога / сем. признаки: базовый: абстрактное ИЛИ базовый: действие

СОБЛЮДАТЬ, = Тщательно оберегать; поддерживать

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = винительный / предлог = без предлога / сем. признаки: базовый: абстрактное

УЧАСТВОВАТЬ, = Принимать участие в чем-л.

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; второй актант: падеж = предложный / предлог = «в» / сем. признаки: базовый: действие

УЧАСТВОВАТЬ, = Иметь свою долю, пай в каком-л. деле, предприятия

(условия реализации) глагол; первый актант: падеж = именительный / сем. признаки: базовый: живое существо; человек = +; второй актант: падеж = предложный / предлог = «в» / сем. признаки: базовый: действие

Рассмотрим теперь, как взаимодействуют семантические признаки и их значения, выделенные для существительных и глаголов в соответствующих словосочетаниях. Возьмем словосочетание *окружить дачу забором*.

У существительного *дача* выделяются два значения (см. выше) — предмет ('строение, дом') и место ('участок земли'). В ряде случаев они выступают нераздельно, синкретично, ср. *поехал на дачу, живет на даче*. Однако в некоторых случаях они явно дифференцируются, ср. *дача сгорела* ('дом') vs. *сажать картошку на даче* ('участок земли').

У глагола *окружить* выделяется четыре ССВ (см. выше). В первом значении этот глагол имеет две валентности, в трех других — три валентности, которые выражаются синтаксически тождественным способом. Из этого следует, что в словосочетании *окружить дачу забором* глагол *окружить* теоретически может выступать в любом из этих трех значений.

Чисто комбинаторно, не принимая во внимание семантическую характеристику глагольных актантов, мы получаем шесть вариантов интерпретации (два у существительного помножить на три у глагола). Благодаря учету значений базового семантического признака у второго и третьего глагольных актантов удастся отсечь два последних значения глагола *окружить* и, следовательно, четыре комбинаторных варианта.

Остается два варианта, обусловленные неоднозначностью существительного. Как нам кажется, оба семантически оправданны. В одном случае речь идет об окружении дома забором, во втором — об огораживании участка земли забором. По-видимому, это тот случай, когда указанные значения существительного *дача* выступают слитно, нерасчлененно, синкретично.

Приложение 3

ПРИМЕРЫ СЕМАНТИЧЕСКОГО РАЗБОРА ЗАКОНОДАТЕЛЬНЫХ ТЕКСТОВ

В данном приложении мы опускаем технические подробности графического, морфологического и синтаксического разборов и показываем только те фрагменты текстов из региональных законодательных актов, неоднозначность или некорректность которых не бросаются в глаза при прочтении их человеком, однако обращают на себя внимание при автоматическом разборе текста. Все примеры взяты из постановлений одного из областных правительств.

1. *О проведении внеплановой проверки Комитет уведомляет проверяемое лицо (организацию) в течение трех рабочих дней со дня издания Приказа о проведении проверки Комитетом и запрашивает у него документы, указанные в Приказе о проведении проверки.*

Комитет уведомляет... лицо аналогично знаменитому примеру *Мать любит дочь* в том, что касается неоднозначности синтаксической структуры (выбора подлежащего и прямого дополнения). Справедливости ради нужно сказать, что в официальных документах всегда используют прямой порядок слов. К тому же в данном примере невозможность инверсии исключается тем, что добавление в скобках слова *организацию* исключает двоякое толкование (*комитет уведомляет организацию*). Тем не менее подобные случаи нередко встречаются в документах, а любая неоднозначность недопустима в официальных текстах. Ср. также: *Орган финансового контроля уведомляет субъект финансового контроля о проведении мероприятия финансового контроля (орган уведомляет субъект).*

2. В следующих примерах вместо выражения *включать в себя* используется просто *включать*. Можно *включить что-то* в список, но нельзя сказать, что список *включает что-то*. Список может *состоять из чего-то* или *включать в себя*. *Постановка и проектирование* не могут ничего *включать*.

2.1. *Постановка животного на учет включает присвоение животному уникального идентификационного номера животного, мечение животного, выдачу документа о постановке животного на учет или внесение соответствующих отметок в паспорт животного.*

2.2 *Основное мероприятие 1. Проектирование и строительство систем тепло-, водоснабжения и водоотведения на территории муниципального образования «Город Псков» включает следующие мероприятия...*

3. *Запретить выход населения на лед водных объектов города Пскова с момента ледостава и до образования прочного ледового покрова.*

Эта фраза создает как минимум две проблемы при автоматическом анализе текста:

- При формальном синтаксическом анализе мы получим *объекты города Пскова*, но при этом непонятно, о каком *населении* идет речь (население какого пункта?). Более того, даже если предположить, что речь идет о населении Пскова, получается, что индивидам, не относящимся к населению Пскова, не запрещается выходить на лед.
- *Население* определяется как ‘совокупность жителей’, хотя на деле запрет относится к отдельным индивидам или их группам, которые могут выйти на лед (прежде всего рыбакам), а не ко всему населению Пскова.

4. *Система управления гражданской обороны составляет совокупность взаимосвязанных органов и пунктов управления всех звеньев, оснащенных системой связи, оповещения и управления.*

Система может состоять из чего-то, а не *составлять* что-то. На самом деле все наоборот: совокупность органов и пунктов составляет систему управления.

Управление, если это не название организации, учреждения (например, *Управление гражданской обороны*), требует зависимого существительного в творительном падеже без предлога (*управление* — чем?). В результате мы должны получить такую фразу: *Система управления гражданской обороной состоит из совокупности органов и пунктов управления...*

К тому же данная фраза не просто тавтологична, а, если так можно сказать, избыточно тавтологична, ср.: *Система **управления**... состоит из органов и пунктов **управления**, оснащенных... **управлением**.*

Подобных примеров в законодательных текстах можно найти немало. Излишне говорить, что законодательные тексты сильно страдают канцеляризмами и тавтологией. Синтаксическая структура предложений нередко допускает множественные прочтения, что затрудняет не только автоматический анализ, но даже обычное понимание носителем языка.

Научное издание

*КУЗНЕЦОВ Сергей Александрович, СКРЕБЦОВА Татьяна Георгиевна,
СУВОРОВ Сергей Геннадиевич, КЛЕМЕНТЬЕВА Анна Владимировна*

ЛИНГВИСТИЧЕСКИЙ АНАЛИЗАТОР

Преобразование текста в метаязыковую структуру данных

Редактор *Е. П. Парфенова*

Корректор *А. С. Яшина*

Компьютерная верстка *Е. М. Воронковой*

Обложка *Е. Р. Куныгина*

Подписано в печать 11.03.19. Формат 60×90 1/16.

Усл. печ. л. 14,8. Плановый тираж 500 экз. (1-й завод — 150 экз.). Заказ № .

Издательство Санкт-Петербургского университета.

199004, С.-Петербург, В.О., 6-я линия, 11. Тел./факс +7(812)328-44-22

publishing@spbu.ru



publishing.spbu.ru

Типография Издательства СПбГУ. 199034, С.-Петербург, Менделеевская линия, д. 5.

Книги **Издательства СПбГУ** можно

ЗАКАЗАТЬ

на сайте издательства: **publishing.spbu.ru**

в интернет-магазинах: **ozon.ru**; **bookvoed.ru**; **URSS.ru**

ПРИБРЕСТИ

в книжных магазинах СПбГУ ***Дом университетской книги***

Менделеевская линия, д. 5

6-я линия В. О., д. 15

Университетская наб., д. 11

А также в магазинах

Санкт-Петербурга: Сеть книжных магазинов «Буквоед»

Санкт-Петербургский Дом книги, Невский пр., д. 28

«Подписные издания», Литейный пр., д. 57

Москвы:

«Библио-Глобус», ул. Мясницкая, д. 6/3

«Фаланстер», М. Гнезниковский пер., д. 12/27

Перми:

«Пиотровский», ул. Ленина, д. 54